

# Distributed Detection over Adaptive Networks: Refined Asymptotics and the Role of Connectivity

Vincenzo Matta, Paolo Braca, Stefano Marano, Ali H. Sayed

## Abstract

We consider distributed detection problems over adaptive networks, where dispersed agents learn continually from streaming data by means of local interactions. The requirement of adaptation allows the network of detectors to track drifts in the underlying hypothesis. The requirement of cooperation allows each agent to deliver a performance superior to what would be obtained if it were acting individually. The simultaneous requirements of adaptation and cooperation are achieved by employing diffusion algorithms with constant step-size  $\mu$ . In [1], [2] some main features of adaptive distributed detection were revealed. By resorting to large deviations analysis, it was established that the Type-I and Type-II error probabilities of all agents vanish exponentially as functions of  $1/\mu$ , and that all agents share the same Type-I and Type-II error exponents. However, numerical evidences presented in [1], [2] showed that the theory of large deviations does not capture the fundamental impact of network connectivity on performance, and that additional tools and efforts are required to obtain accurate predictions for the error probabilities. This work addresses these open issues and extends the results of [1], [2] in several directions. By conducting a refined asymptotic analysis based on the mathematical framework of exact asymptotics, we arrive at a revealing and powerful understanding of the *universal behavior of distributed detection over adaptive networks*: as functions of  $1/\mu$ , the error (log-)probability curves corresponding

V. Matta and S. Marano are with DIEM, University of Salerno, via Giovanni Paolo II 132, I-84084, Fisciano (SA), Italy (e-mail: vmatta@unisa.it; marano@unisa.it).

P. Braca is with NATO STO Centre for Maritime Research and Experimentation, La Spezia, Italy (e-mail: paolo.braca@cmre.nato.int).

A. H. Sayed is with the Electrical Engineering Department, University of California, Los Angeles, CA 90095 USA (e-mail: sayed@ee.ucla.edu).

A short version of this work appeared in the conference publication [3].

to different agents stay nearly-parallel to each other (as already discovered in [1], [2]), however, these curves are ordered following a criterion reflecting the degree of connectivity of each agent. Depending on the combination weights, the more connected an agent is, the lower its error probability curve will be. The analysis provides explicit analytical formulas for the detection error probabilities and these expressions are also verified by means of extensive simulations. We further enlarge the reference setting from the case of doubly-stochastic combination matrices considered in [1], [2], to the more general and demanding setting of right-stochastic combination matrices; this extension poses new and interesting questions in terms of the interplay between the network topology, the combination weights, and the inference performance. The potential of the proposed methods is illustrated by application of the results to canonical detection problems, to typical network topologies, for both doubly-stochastic and right-stochastic combination matrices. Interesting and somehow unexpected behaviors emerge, and the lesson learned is that *connectivity matters*.

### Index Terms

Distributed detection, adaptive network, diffusion strategy, large deviations analysis, exact asymptotics.

## I. INTRODUCTION AND MOTIVATION

Adaptive networks consist of spatially dispersed agents, learning continually from streaming data by means of distributed processing algorithms. With a proper form of cooperation, each agent is able to deliver an inference performance superior to what would be obtained if it were acting individually. When called to operate in complex and dynamical scenarios (e.g., in the presence of drifts in the statistical conditions, in the environmental conditions, and in the network topology, among other possibilities), the agents must be further endowed with strong adaptive capabilities, in order to respond in real-time to these variations.

Several useful distributed inference solutions are available that meet these requirements. Particularly relevant to our setting are the diffusion implementations based on adaptive algorithms with constant step-size [4]–[7]. The use of a constant step-size (as opposed to the decaying step-size employed in the case of consensus algorithms [8]–[19]) is key to enable continuous learning, e.g., to meet the fundamental requirement of tracking. The interplay between adaptation and learning is critical for guaranteeing the successful network operation and to produce reliable inference.

Such an interplay has been examined rather deeply in the framework of *estimation* problems, while less attention has been devoted to *detection* problems. Distributed detection over adaptive networks can be succinctly described as follows. A network of dispersed agents, linked together by a given topology, monitors a certain physical phenomenon. As time elapses, the agents gather from the environment streaming data, whose statistical properties depend upon an unknown state of nature, formally represented by a pair of hypotheses, say,  $\mathcal{H}_0$  and  $\mathcal{H}_1$ . At *each* time instant, *each* individual agent must produce a decision inferring the current state of nature, using its own available observations and the local exchange of information obtained from consultation with its neighbors. Due to the requirement of adaptation, the agents must be able to react promptly to drifts in the current state of nature, while guaranteeing an adequate performance level (i.e., low detection error probabilities) at the steady-state, namely, when a given hypothesis is in force for sufficiently long time.

With reference to the above setting, in the recent works [1], [2] several interesting results were derived that allowed the authors to establish fundamental scaling laws for adaptive distributed detection by multi-agent networks. In particular, it was shown there that for diffusion strategies with constant step-size  $\mu$ , the steady-state error probabilities exhibit an exponential decay, as  $\mu \rightarrow 0$ , as functions of  $1/\mu$ . By resorting to a detailed large deviations analysis, the exact scaling law was fully characterized in [1], [2] in terms of a rate function  $\Phi(\gamma)$  that can be evaluated analytically. Notably, the results of [1], [2] showed that all network agents share the same error exponent so that their error probabilities are asymptotically equivalent to the leading exponential order as  $\mu \rightarrow 0$ . This is a remarkable feature of adaptive distributed detection solutions. Readers may consult the introductory remarks of [1], [2], which contain a detailed and motivated summary of these results.

One known limitation of large deviations analysis resides in the fact that it focuses only on the leading exponential order, thus neglecting all sub-exponential terms. The practical implication of this fact can be easily illustrated by a simple example. Assume network agents 1 and 2 exhibit asymptotic error probabilities  $P_1$  and  $P_2$  of the form:

$$P_1 = e^{-\frac{1}{\mu}}, \quad P_2 = 2e^{-\frac{1}{\mu}} = e^{-\frac{1}{\mu}[1+o(1)]}, \quad (1)$$

where  $o(1)$  stands for any correction such that  $o(1) \rightarrow 0$  as  $\mu \rightarrow 0$ . These two probabilities have the same error exponent multiplying  $-1/\mu$  (and is equal to one), but the error probability at

agent 2 is always twice that of agent 1. This is because the factor 2 is a sub-exponential term.

The above consideration is particularly relevant in the context of adaptive distributed detection since numerical experiments reported in [1], [2] show that, depending on the particular network structure, different agents have different error probabilities, while exhibiting the same scaling law to the leading exponential order. Even more remarkably, the experimental results of [1], [2] suggested that the differences in the error probabilities appear to be related to the degrees of connectivity of the various agents. Accordingly, the main purpose of this paper is to perform a refined asymptotic analysis overcoming the limitations of large deviations, in order to capture the fundamental dependence of the error probabilities at different agents upon the specific network connectivity. In this way, we will be able to show that the partial empirical evidences in [1], [2] are in fact representative of a *universal behavior of distributed detection over adaptive networks*.

Another relevant contribution provided in this work is that, differently from what was done in [1], [2], we shall not limit the analysis to doubly-stochastic combination matrices and will consider instead the broader class of right-stochastic matrices. This latter class is relevant in practical applications. Indeed, while doubly-stochastic matrices might look preferable in detection applications in view of the asymptotic equipartition of the combination weights, they are not always realizable in practice. This happens, for instance, due to physical limitations in the topology and/or communication constraints. This generalization in the nature of the combination policy poses new and interesting questions in terms of the interplay between the network topology, the combination weights, and the inference performance.

In summary, as a result of the refined analysis presented in this work, we will not only be able to evaluate in quantitative terms how network connectivity reflects on the performance of different agents, but we will also be able to clarify how different network structures (i.e., different combination matrices and network topologies) globally impact the network performance.

### A. Related Work

Distributed detection is a classical topic, and the pertinent literature is therefore extensive. Any set of references would be by-no-means exhaustive, so that we refer the reader to [20]–[31] as fundamental entry points on the subject. In this work we consider fully-decentralized detection problems, i.e., fully-flat network architectures without fusion center, where the agents are only allowed to interact locally. Several recent works address this scenario and, in particular, solutions

based on decentralized consensus strategies with decaying step-size have been successfully proposed in [8]–[13], and the detection performance of these algorithms has been accurately characterized in different asymptotic frameworks [8]–[13]. However, as already observed in the introduction, a distinguishing feature of our work resides in the emphasis on *adaptive* solutions. To enable adaptation, it has been shown that diffusion algorithms with constant step-size are superior to consensus implementations due to an inherent asymmetry in the consensus update that has been shown to be a source of potential instability in the consensus dynamics when constant step-sizes are used [4]–[7]. Several performance results are already available for diffusion strategies in connection to their mean-square-error (MSE) estimation behavior [4]–[7]. The corresponding results for detection applications are relatively limited. In [32] the problem of using diffusion algorithms for detection purposes has been considered, with reference to a Gaussian problem. More recently, the general problem of distributed detection over adaptive networks has been posed in [1], [2]. By means of a large deviations analysis, the Type-I and Type-II error exponents, for doubly-stochastic combination matrices, were characterized in closed-form in these works. It was shown that the detection performance of the network solution is equivalent to the fully-connected solution to the first-leading order in the exponent. However, numerical and simulation results in [1], [2] showed that the network connectivity and the overall structure of the combination weights do matter, and they influence the performance of the individual agents. Analytical formulas for accurate evaluation of the detection error probabilities, and for elucidating the relationship between the network structure and the expected performance, are currently missing. Filling this important gap is the main theme of this work.

### B. Summary of Main Result

In order to introduce the main result, we refer to the steady-state (i.e., as time goes to infinity) output of the diffusion algorithm with constant step-size  $\mu$ . The steady-state output at the  $k$ -th agent is denoted by  $\mathbf{y}_{k,\mu}^*$ . It was shown in [1], [2] that, for small step-sizes,  $\mathbf{y}_{k,\mu}^*$  tends to concentrate in the close proximity of the expected value of the local statistics, denoted by  $\mathbb{E}[\mathbf{x}]$ . In order to characterize the error probabilities, we shall evaluate the probability that  $\mathbf{y}_{k,\mu}^*$  deviates significantly from this expected behavior. Without loss of generality, we shall focus on the following probability:

$$\mathbb{P}[\mathbf{y}_{k,\mu}^* > \gamma], \quad \gamma > \mathbb{E}[\mathbf{x}]. \quad (2)$$

With reference to the case of doubly-stochastic combination matrices, it was established in [1], [2] that

$$\lim_{\mu \rightarrow 0} \mu \ln \mathbb{P}[\mathbf{y}_{k,\mu}^* > \gamma] = -\Phi(\gamma), \quad (3)$$

where  $\Phi(\gamma)$  is the so-called *rate function*. Equivalently, we can rewrite (3) as

$$\mathbb{P}[\mathbf{y}_{k,\mu}^* > \gamma] = e^{-\frac{1}{\mu} [\Phi(\gamma) + o(1)]}. \quad (4)$$

This last form of representation highlights the fact that sub-exponential terms are neglected by large deviations analysis.

A refined analysis can be pursued by seeking an asymptotic approximation,  $\mathcal{P}_{k,\mu}(\gamma)$ , that ensures the much stronger conclusion:

$$\mathbb{P}[\mathbf{y}_{k,\mu}^* > \gamma] = \mathcal{P}_{k,\mu}(\gamma)[1 + o(1)], \quad (5)$$

i.e.,

$$\lim_{\mu \rightarrow 0} \frac{\mathbb{P}[\mathbf{y}_{k,\mu}^* > \gamma]}{\mathcal{P}_{k,\mu}(\gamma)} = 1. \quad (6)$$

This framework is commonly referred to as *exact asymptotics*, and has been originally studied in [33] with reference to the simplest case of normalized sums of independent and identically distributed (i.i.d.) random variables — see also [34], [35]. In the following, we shall often write

$$\mathbb{P}[\mathbf{y}_{k,\mu}^* > \gamma] \sim \mathcal{P}_{k,\mu}(\gamma) \quad (7)$$

to denote an asymptotic equivalence of the kind (6). It is immediately seen that (6) implies

$$\lim_{\mu \rightarrow 0} \mu \ln \mathbb{P}[\mathbf{y}_{k,\mu}^* > \gamma] = \lim_{\mu \rightarrow 0} \mu \ln \mathcal{P}_{k,\mu}(\gamma), \quad (8)$$

which means that any exact asymptotic  $\mathcal{P}_{k,\mu}(\gamma)$  is able to reflect the leading exponential term.

The main result established in this paper can now be formally stated as follows:

$$\boxed{\mathcal{P}_{k,\mu}(\gamma) = \sqrt{\frac{\mu}{2\pi\theta_\gamma^2 \phi''(\theta_\gamma)}} e^{-\frac{1}{\mu} [\Phi(\gamma) + \epsilon_{k,\mu}(\theta_\gamma)]}} \quad (9)$$

The main quantities necessary to evaluate the asymptotic approximation  $\mathcal{P}_{k,\mu}(\gamma)$  in the above expression are now briefly introduced. The function  $\phi(t)$  will be described in closed-form, and depends on the underlying statistical model through the moment generating function of the local

statistics, and on the network topology only through the *limiting* vector of combination weights (aka the Perron eigenvector, see, e.g., [7]). The quantity  $\theta_\gamma$  is the solution to the equation:

$$\phi'(\theta_\gamma) = \gamma. \quad (10)$$

The function  $\Phi(\gamma)$ , which was the main object of study in [1], [2], is the so-called *rate function*, and is computed directly from  $\phi(t)$ , namely, it is the Fenchel-Legendre transform of  $\phi(t)$  [34], [35].

Finally, the (sub-exponential) correction term  $\epsilon_{k,\mu}(\theta_\gamma)$  depends on the underlying statistical model through the moment generating function of the local statistics, and on the network topology through the *actual* (not only the limiting) network combination weights. This correction satisfies:

$$\epsilon_{k,\mu}(\theta_\gamma) \xrightarrow{\mu \rightarrow 0} 0. \quad (11)$$

In particular, the ratio  $\epsilon_{k,\mu}(\theta_\gamma)/\mu$  stays bounded as  $\mu \rightarrow 0$ . Loosely speaking, this means that the overall correction appearing at the exponent in (9) plays, asymptotically, the role of a *constant* correction.

Despite its apparent complexity, Eq. (9) possesses a well defined structure, revealing important connections with the physical behavior of the adaptive distributed system under consideration. Let us elucidate some of these features. We start by the leading order in the exponent. It is easy to see that:

$$\begin{aligned} \mu \ln \mathcal{P}_{k,\mu}(\gamma) &= -\Phi(\gamma) - \epsilon_{k,\mu}(\theta_\gamma) + \\ &\quad \frac{\mu}{2} \ln \mu - \frac{\mu}{2} \ln [2\pi\theta_\gamma^2 \phi''(\theta_\gamma)] \\ &\xrightarrow{\mu \rightarrow 0} -\Phi(\gamma), \end{aligned} \quad (12)$$

which easily follows from (11). This means that the approximation  $\mathcal{P}_{k,\mu}$  in (9) can be regarded as

$$\mathcal{P}_{k,\mu}(\gamma) = e^{-\frac{1}{\mu}[\Phi(\gamma) + o(1)]}. \quad (13)$$

This is consistent with result (3), which was established in [1], [2] for the case of doubly-stochastic combination matrices. The terms of order  $o(1)$  in (13) collect all the sub-exponential corrections that appear in (9). They can be separated into two categories. The first correction is the term  $\sqrt{\frac{\mu}{2\pi\theta_\gamma^2 \phi''(\theta_\gamma)}}$  in (9). This term is a typical sub-exponential refinement arising in the framework of exact asymptotics, and is a consequence of a local Central Limit Theorem

(see [33]–[35]). Observe that this correction, which is related to the network topology only through the Perron eigenvector, is *independent* of the agent index  $k$ , and is therefore applicable to all agents. In contrast, the second correction  $\epsilon_{k,\mu}(\theta_\gamma)$  depends on the agent index  $k$  and, as it will be detailed in the exact statement of the main theorems, takes into account the entire network topology and combination weights.

The above considerations lead to the important conclusion that (9) provides a detailed and revealing assessment of the *universal behavior of distributed detection over adaptive networks*: as functions of  $1/\mu$ , the error (log-)probability curves corresponding to different agents not only stay nearly-parallel to each other, but they are also ordered following a criterion dictated by the correction term  $\epsilon_{k,\mu}(\theta_\gamma)$ . As we shall see later — see Fig. 1 for an example, this criterion reflects the degree of connectivity of each agent. Depending on the combination weights, the more connected an agent is, the lower its error probability will be, and the correction term  $\epsilon_{k,\mu}(\theta_\gamma)$  is sufficiently rich to capture this behavior.

**Notation.** We use boldface letters to denote random variables, and normal font letters for their realizations. Capital letters refer to matrices, small letters to both vectors and scalars. Sometimes we violate this latter convention, for instance, we denote the total number of sensors by  $S$ . The symbols  $\mathbb{P}$  and  $\mathbb{E}$  are used to denote the probability and expectation operators, respectively. The notation  $\mathbb{P}_h$  and  $\mathbb{E}_h$ , with  $h = 0, 1$ , means that the pertinent statistical distribution corresponds to hypothesis  $\mathcal{H}_0$  or  $\mathcal{H}_1$ .

## II. PROBLEM FORMULATION

We consider a sensor network that collects observations about a physical phenomenon of interest. Data are assumed to be spatially and temporally independent and identically distributed. From the observation measured at time  $n$ , the  $k$ -th agent computes its local statistic (the observation itself, or a suitable function thereof), which is denoted by  $\mathbf{x}_k(n)$ ,  $k = 1, 2, \dots, S$ . The mean and variance of  $\mathbf{x}_k(n)$  will be denoted by  $\mathbb{E}[\mathbf{x}]$  and  $\sigma_x^2$ , respectively. To avoid trivialities, throughout the paper it is assumed that the random variable  $\mathbf{x}_k(n)$  is non-degenerate.

### A. Diffusion Strategy

Following the framework developed in [1], [2], we focus on the class of diffusion strategies for adaptation over networks [6], [7], [32], and in particular on the ATC (Adapt-Then-Combine)



implementation due to some inherent advantages in terms of a slightly improved mean-square-error performance relative to other forms [6]. Extensions to other diffusion implementations, as well as to consensus implementations, are certainly possible. In the ATC algorithm, each node  $k$  updates its state from  $\mathbf{y}_k(n-1)$  to  $\mathbf{y}_k(n)$  through local cooperation with its neighbors as follows:

$$\mathbf{v}_k(n) = \mathbf{y}_k(n-1) + \mu[\mathbf{x}_k(n) - \mathbf{y}_k(n-1)], \quad (14)$$

$$\mathbf{y}_k(n) = \sum_{\ell=1}^S a_{k,\ell} \mathbf{v}_\ell(n), \quad (15)$$

where  $0 < \mu \ll 1$  is a small step-size parameter. It is seen that node  $k$  first uses its locally available statistic  $\mathbf{x}_k(n)$ , to update its state from  $\mathbf{y}_k(n-1)$  to an intermediate value  $\mathbf{v}_k(n)$ . The other network agents simultaneously perform similar updates using their local statistics. Subsequently, node  $k$  aggregates the intermediate states of its neighbors using nonnegative convex combination weights  $\{a_{k,\ell}\}$  that add up to one. Again, all other network agents perform a similar calculation. Collecting the combination weights into a square matrix  $A = [a_{k,\ell}]$ , then  $A$  is a right-stochastic matrix, namely, the entries on each row add up to one. Formally:

$$a_{k,\ell} \geq 0, \quad A\mathbf{1} = \mathbf{1}, \quad (16)$$

with  $\mathbf{1}$  being a column-vector with all entries equal to 1. We denote the  $n$ -th power of  $A$  by

$$B_n = [b_{k,\ell}(n)] \triangleq A^n. \quad (17)$$

Throughout this article, we assume that  $A$  has second largest eigenvalue magnitude strictly less than one, which yields [15], [36]:

$$\boxed{b_{k,\ell}(n) \xrightarrow{n \rightarrow \infty} p_\ell \Leftrightarrow B_n \xrightarrow{n \rightarrow \infty} \mathbf{1}p} \quad (18)$$

where the limiting combination weights  $\{p_\ell\}$  satisfy, for all  $\ell = 1, 2, \dots, S$ :

$$\boxed{pA = p, \quad p_\ell > 0, \quad \sum_{\ell=1}^S p_\ell = 1} \quad (19)$$

and the row vector  $p = [p_1, p_2, \dots, p_S]$  is usually referred to as the Perron eigenvector of  $A$  — see, e.g., [7]. We remark that the condition on  $A$  is automatically satisfied by network topologies that are strongly-connected [7], i.e., when there is always a path with nonzero weights between

any pair of nodes, and at least one node in the network has a self-loop ( $a_{k,k} > 0$  for some agent  $k$ ).

### B. Steady-State Distribution

In order to design and characterize an inference system based upon the sensor output  $\mathbf{y}_k(n)$ , knowledge of the distribution of  $\mathbf{y}_k(n)$  is crucial. However this knowledge is seldom available, except for very special cases (e.g., Gaussian observations). A common and well-established approach in the adaptation literature [6], [37] to address this difficulty is to focus on *i*) the steady-state properties (as  $n \rightarrow \infty$ ), and *ii*) the small step-size regime ( $\mu \rightarrow 0$ ). Accordingly, throughout the paper, the term *steady-state* refers to the limit as the time-index  $n$  goes to infinity, while the term *asymptotic* refers to the slow adaptation regime where  $\mu \rightarrow 0$ .

We start by considering the steady-state behavior of  $\mathbf{y}_k(n)$  for a given step-size  $\mu$ . To this aim, it is useful to recast the pair of equations given by (14) and (15) into the following single equation, which is obtained by straightforward algebra:

$$\begin{aligned} \mathbf{y}_k(n) = & \underbrace{(1 - \mu)^n \sum_{\ell=1}^S b_{k,\ell}(n) \mathbf{y}_\ell(0)}_{\text{transient term}} + \\ & \sum_{i=1}^n \sum_{\ell=1}^S \mu (1 - \mu)^{i-1} b_{k,\ell}(i) \mathbf{x}_\ell(n - i + 1). \end{aligned} \quad (20)$$

Since the random variables  $\mathbf{x}_k(n)$  are i.i.d. across time, and since we shall be only concerned with the distribution of partial sums involving these terms, it is convenient to define the following random variable:

$$\mathbf{y}_k^*(n) \triangleq \sum_{i=1}^n \sum_{\ell=1}^S \mu (1 - \mu)^{i-1} b_{k,\ell}(i) \mathbf{x}_\ell(i), \quad (21)$$

which *shares the same distribution of* the second term on the RHS of (20) — see also the discussion in [1], [2]. Formally:

$$\mathbf{y}_k^*(n) \stackrel{d}{=} \sum_{i=1}^n \sum_{\ell=1}^S \mu (1 - \mu)^{i-1} b_{k,\ell}(i) \mathbf{x}_\ell(n - i + 1), \quad (22)$$

where the notation  $\stackrel{d}{=}$  denotes equality *in distribution*.

In Theorem 1 of [1], [2], the existence of a steady-state random variable characterizing the diffusion output has been established. This can be summarized by the following statement (the symbol  $\rightsquigarrow$  means convergence in distribution):

$$\boxed{\mathbf{y}_k(n) \xrightarrow{n \rightarrow \infty} \mathbf{y}_{k,\mu}^* \triangleq \sum_{i=1}^{\infty} \sum_{\ell=1}^S \mu (1 - \mu)^{i-1} b_{k,\ell}(i) \mathbf{x}_\ell(i)} \quad (23)$$

where the first two moments of  $\mathbf{y}_{k,\mu}^*$  are given by

$$\mathbb{E}[\mathbf{y}_{k,\mu}^*] = \mathbb{E}[\mathbf{x}], \quad (24)$$

$$\text{VAR}[\mathbf{y}_{k,\mu}^*] = \sigma_x^2 \sum_{i=1}^{\infty} \sum_{\ell=1}^S \mu^2 (1 - \mu)^{2(i-1)} b_{k,\ell}^2(i). \quad (25)$$

Actually, the proof in [1], [2] focused on the case of doubly-stochastic combination matrices, but it is immediate to verify that the same argument holds for right-stochastic combination matrices and leads to (23)–(25).

### C. The Inference Problem

In the distributed detection formulation, each agent in the network must perform a binary hypothesis test, in an adaptive and fully decentralized manner. In this setting, the agents in the network continually collect an increasing amount of streaming data, whose statistical properties depend upon an *unknown* binary state of nature, which is represented by a pair of hypotheses, say,  $\mathcal{H}_0$  and  $\mathcal{H}_1$ . The statistics  $\mathbf{x}_k(n)$  are spatially and temporally i.i.d., *conditioned* on the hypothesis that gives rise to them. In what follows, we shall always assume that  $\mathbb{E}_0[\mathbf{x}] \neq \mathbb{E}_1[\mathbf{x}]$ . To get some intuition about the meaning of this condition, consider that, when the local statistic  $\mathbf{x}$  is chosen as the log-likelihood of the local observation, the inequality  $\mathbb{E}_0[\mathbf{x}] \neq \mathbb{E}_1[\mathbf{x}]$  is simply a way to state that the detection problem is *identifiable*, namely, that it is not singular [38]. Furthermore, and without loss of generality, we assume that:

$$\mathbb{E}_0[\mathbf{x}] < \mathbb{E}_1[\mathbf{x}]. \quad (26)$$

At time  $n$ , the  $k$ -th sensor needs to produce a decision about the state of nature, based upon its state value  $\mathbf{y}_k(n)$ . As discussed in [1], [2], the computation of  $\mathbf{y}_k(n)$  via algorithm (14) and (15) is motivated by the fact that this implementation essentially results in a value  $\mathbf{y}_k(n)$  that corresponds to a weighted average of the local statistics  $\mathbf{x}_k(n)$ . Such additive constructions for

the state variable are not only convenient from an implementation point of view, but they are also meaningful from a theoretical standpoint. Indeed, in the classical, centralized and non-adaptive theory with i.i.d. data (where the optimal detection statistic is the sum of the log-likelihoods), as well as in more general frameworks such as locally optimum detection or universal hypothesis testing [38], [39], using additive detection statistics is the best choice. Moreover, the decision regions employed by the detector are often in the form of single-threshold rules, which, for a variety of detection problems, exhibit asymptotic optimality properties also in our adaptive and distributed setting, as already shown in [1], [2]. Motivated by these considerations, in this work the test implemented by agent  $k$  at time  $n$  is therefore chosen to be of the form:

$$\mathbf{y}_k(n) \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\leq}} \gamma. \quad (27)$$

The performance of the test is typically expressed in terms of the Type-I (choose  $\mathcal{H}_1$  when  $\mathcal{H}_0$  is true) and Type-II (choose  $\mathcal{H}_0$  when  $\mathcal{H}_1$  is true) error probabilities. With reference to the steady-state performance, the Type-I and Type-II error probabilities are defined as, respectively:

$$\alpha_{k,\mu} \triangleq \mathbb{P}_0[\mathbf{y}_{k,\mu}^* > \gamma], \quad \beta_{k,\mu} \triangleq \mathbb{P}_1[\mathbf{y}_{k,\mu}^* \leq \gamma]. \quad (28)$$

### III. MAIN THEOREMS

We now introduce the basic quantities necessary to characterize the system performance. As we shall see, it is important to understand the behavior of the Logarithmic Moment Generating Function (LMGF) of the steady-state variable  $\mathbf{y}_{k,\mu}^*$ . In Appendix A we state and prove a theorem (Theorem 1) that establishes several useful properties of the LMGF, of its derivatives, and of its limiting behavior as  $\mu \rightarrow 0$ . In the following description we shall mention and use some of these properties, referring the reader to Appendix A for a more detailed explanation.

- *Local LMGF.*

The LMGF of the local statistics  $\mathbf{x}_k(n)$  is defined as:

$$\psi(t) \triangleq \ln \mathbb{E}[e^{t\mathbf{x}_k(n)}], \quad (29)$$

and is independent of  $k$ . A fundamental role in our results will be played by the following averaged version of the LMGF  $\psi(t)$ , which has already been used in [1], [2]:

$$\omega(t) \triangleq \int_0^t \frac{\psi(\tau)}{\tau} d\tau. \quad (30)$$

- *Steady-state LMGF.*

The LMGF of the steady-state variable  $\mathbf{y}_{k,\mu}^*$  is defined as:

$$\phi_{k,\mu}(t) \triangleq \ln \mathbb{E}[e^{t\mathbf{y}_{k,\mu}^*}], \quad (31)$$

and admits the following representation [Theorem 1, Eq. (88)]:

$$\phi_{k,\mu}(t) = \sum_{i=1}^{\infty} \sum_{\ell=1}^S \psi(\mu(1-\mu)^{i-1} b_{k,\ell}(i)t). \quad (32)$$

- *Limiting properties of the steady-state LMGF.*

We will be primarily concerned with the limiting normalized LMGF:

$$\phi(t) \triangleq \lim_{\mu \rightarrow 0} \mu \phi_{k,\mu}(t/\mu), \quad (33)$$

where the above limit exists, and is given by [Theorem 1, Eq. (89)]:

$$\boxed{\phi(t) = \sum_{\ell=1}^S \omega(p_{\ell}t) = \sum_{\ell=1}^S \int_0^{p_{\ell}t} \frac{\psi(\tau)}{\tau} d\tau} \quad (34)$$

Equation (34) emphasizes that  $\phi(t)$  depends on the underlying statistical model through the LMGF  $\psi(t)$  of the local statistics  $\mathbf{x}_k(n)$ , and on the network topology only through the Perron eigenvector  $p$ , i.e., on the limiting combination weights. For doubly-stochastic matrices we have  $p_{\ell} = 1/S$ , so that the above formula gives

$$\phi(t) = S\omega(t/S) = S \int_0^{t/S} \frac{\psi(\tau)}{\tau} d\tau, \quad (35)$$

which is consistent with the results of [1], [2].

It is of interest to consider an alternative representation for (34). To this aim, we introduce the LMGF of the averaged random variable  $\sum_{\ell=1}^S p_{\ell} \mathbf{x}_{\ell}(n)$ , namely:

$$\bar{\psi}(t) \triangleq \sum_{\ell=1}^S \psi(p_{\ell}t), \quad (36)$$

and by straightforward calculations we get from (34):

$$\boxed{\phi(t) = \int_0^t \frac{\bar{\psi}(\tau)}{\tau} d\tau} \quad (37)$$

- *Property of the LMGF derivatives.*

In Theorem 1, part *ii*), we establish that the derivatives of  $\phi_{k,\mu}(t)$  and of the limiting

function  $\phi(t)$  can be evaluated by interchanging the differential and limit operators, yielding, in particular [Theorem 1, Eq. (90)]:

$$\begin{aligned} \phi'_{k,\mu}(t) &= \mu \sum_{i=1}^{\infty} \sum_{\ell=1}^S (1-\mu)^{i-1} b_{k,\ell}(i) \times \\ &\quad \psi'(\mu(1-\mu)^{i-1} b_{k,\ell}(i)t), \end{aligned} \quad (38)$$

and [Theorem 1, Eq. (91)]

$$\lim_{\mu \rightarrow 0} \phi'_{k,\mu}(t/\mu) = \phi'(t) = \frac{1}{t} \sum_{\ell=1}^S \psi(p_{\ell}t). \quad (39)$$

- *Convergence errors.*

As we shall see, the convergence errors corresponding to (33) and (39) will play an important role in characterizing the error probabilities. The following rates of convergence [Theorem 1, Eq. (93)] hold:

$$\phi(t) - \mu \phi_{k,\mu}(t/\mu) = \mathcal{O}(\mu) \quad (40)$$

$$\phi'(t) - \phi'_{k,\mu}(t/\mu) = \mathcal{O}(\mu), \quad (41)$$

where the notation  $f_{\mu} = \mathcal{O}(\mu)$  means that the ratio  $f_{\mu}/\mu$  stays bounded as  $\mu \rightarrow 0$ .

- *Fenchel-Legendre transforms.*

The Fenchel-Legendre transform of the function  $\phi(t)$  is defined by [34], [35]:

$$\Phi(\gamma) \triangleq \sup_{t \in \mathbb{R}} [\gamma t - \phi(t)]. \quad (42)$$

We shall use capital letters to denote Fenchel-Legendre transforms, as done in (42). It is further useful to introduce the domain where  $\Phi(\gamma)$  is finite, namely:

$$\mathcal{D}_{\Phi} = \{\gamma \in \mathbb{R} : \Phi(\gamma) < \infty\}. \quad (43)$$

The notation  $\mathcal{D}_{\Phi}^{\circ}$  adopted in the sequel will denote the interior of the set  $\mathcal{D}_{\Phi}$ .

We can now state our second theorem, which generalizes Theorem 3 from [1], [2] to handle the case of right-stochastic matrices.

**THEOREM 2 (Large deviations of  $y_{k,\mu}^*$  as  $\mu \rightarrow 0$ , for right-stochastic matrices).** *Assume that  $\psi(t) < +\infty$  for all  $t \in \mathbb{R}$ . Then, for all  $k = 1, 2, \dots, S$ :*

i) The steady-state variable  $\mathbf{y}_{k,\mu}^*$  obeys a Large Deviation Principle (LDP) with rate function given by the Fenchel-Legendre transform  $\Phi(\gamma)$  defined by (42), namely, for any Borel measurable region  $\Gamma \in \mathbb{R}$ :

$$\lim_{\mu \rightarrow 0} \mu \ln \mathbb{P}[\mathbf{y}_{k,\mu}^* \in \Gamma] = - \inf_{\gamma \in \Gamma} \Phi(\gamma). \quad (44)$$

ii)  $\phi''(t) > 0$  for all  $t \in \mathbb{R}$ , implying that  $\phi(t)$  is strictly convex. Moreover, the rate function  $\Phi(\gamma)$  is strictly convex in  $\mathcal{D}_{\Phi}^o$ , and attains its unique minimum at  $\gamma = \mathbb{E}[\mathbf{x}]$ , with  $\Phi(\mathbb{E}[\mathbf{x}]) = 0$ .

*Proof:* By Theorem 1, part i), Appendix A, we can conclude that the function  $\phi(t)$  in (33) exists and is finite for all  $t \in \mathbb{R}$ . Since, by definition,  $\phi(t)$  is a normalized limiting LMGF, claim i) of the present theorem follows by application of the Gärtner-Ellis Theorem [34], [35].

Claim ii) is a direct extension to the case of right-stochastic combination matrices of the results proved for doubly-stochastic combination matrices in Appendix C of [1].

□

According to property i) in Theorem 2, the steady-state random variable  $\mathbf{y}_{k,\mu}^*$  obeys a LDP with rate function  $\Phi(\gamma)$ . In view of the convexity properties of  $\Phi(\gamma)$  stated in ii), for any  $\gamma \in \mathcal{D}_{\Phi}^o$ , with  $\gamma > \mathbb{E}[\mathbf{x}]$ , the infimum over an interval of the form  $(\gamma, \infty)$  always lies on the boundary point  $\gamma$ . The same conclusion is reached for the infimum over an interval of the form  $(-\infty, \gamma)$  when  $\gamma < \mathbb{E}[\mathbf{x}]$ . Formally, for any  $\gamma \in \mathcal{D}_{\Phi}^o$ , we have:

$$\lim_{\mu \rightarrow 0} \mu \ln \mathbb{P}[\mathbf{y}_{k,\mu}^* > \gamma] = -\Phi(\gamma), \quad \gamma > \mathbb{E}[\mathbf{x}] \quad (45)$$

$$\lim_{\mu \rightarrow 0} \mu \ln \mathbb{P}[\mathbf{y}_{k,\mu}^* \leq \gamma] = -\Phi(\gamma), \quad \gamma < \mathbb{E}[\mathbf{x}]. \quad (46)$$

Note also that, according to Theorem 2, part ii), the choice  $\gamma = \mathbb{E}[\mathbf{x}]$  yields  $\Phi(\gamma) = \Phi(\mathbb{E}[\mathbf{x}]) = 0$ , that is, a null exponent. This is the uninteresting case where the error probability does *not* vanish exponentially. Accordingly, this situation will be ruled out from the forthcoming analysis. Relationships (45) and (46), along with the latter observation, have an immediate implication as regards the choice of the detection threshold. Indeed, in the light of (28), one must ensure that

$$\mathbb{E}_0[\mathbf{x}] < \gamma < \mathbb{E}_1[\mathbf{x}], \quad (47)$$

in order to guarantee the exponential decay of both error probabilities  $\alpha_{k,\mu}$  and  $\beta_{k,\mu}$ .

The exact asymptotics of  $\mathbf{y}_{k,\mu}^*$  are now characterized in the following theorem.

**THEOREM 3 (Exact asymptotics of  $\mathbf{y}_{k,\mu}^*$  as  $\mu \rightarrow 0$ ).** *Assume that  $\mathbf{x}_k(n)$  is not lattice, and that  $\psi(t) < +\infty$  for all  $t \in \mathbb{R}$ . Let  $\gamma \in \mathcal{D}_\Phi^o$ , with  $\gamma > \mathbb{E}[\mathbf{x}]$ , and let  $\theta_\gamma$  be the unique solution to the stationary equation:*

$$\phi'(\theta_\gamma) = \gamma. \quad (48)$$

Then,  $\theta_\gamma > 0$  and, for  $k = 1, 2, \dots, S$ :

$$\mathbb{P}[\mathbf{y}_{k,\mu}^* > \gamma] \sim \mathcal{P}_{k,\mu}(\gamma), \quad (49)$$

where

$$\mathcal{P}_{k,\mu}(\gamma) \triangleq \sqrt{\frac{\mu}{2\pi\theta_\gamma^2 \phi''(\theta_\gamma)}} e^{-\frac{1}{\mu} [\Phi(\gamma) + \epsilon_{k,\mu}(\theta_\gamma)]} \quad (50)$$

with

$$\epsilon_{k,\mu}(t) = \phi(t) - \mu\phi_{k,\mu}(t/\mu). \quad (51a)$$

The correction term  $\epsilon_{k,\mu}(t)$  can be refined to:

$$\epsilon_{k,\mu}(t) = [\phi(t) - \mu\phi_{k,\mu}(t/\mu)] + \frac{[\phi'(t) - \phi'_{k,\mu}(t/\mu)]^2}{2\phi''(t)}. \quad (51b)$$

*Proof:* See Appendix B. □

REMARK I. The key ingredients to computing  $\mathcal{P}_{k,\mu}(\gamma)$  in (50) are the moment generating function  $\psi(t)$  of the local statistics  $\mathbf{x}_k(n)$ , and the combination matrix  $A$ . Indeed, the quantities  $\Phi(\gamma)$  and  $\theta_\gamma$  depend on the function  $\phi(t)$ , which in turn depends on  $\psi(t)$  and on the *limiting* combination weights  $p_\ell$ . The correction term  $\epsilon_{k,\mu}(t)$  (in both versions) depends on  $\psi(t)$  and on the *actual* combination weights  $b_{k,\ell}(i)$ . □

REMARK II. It is useful to comment on the reason for reporting two expressions for  $\epsilon_{k,\mu}(t)$ . First, note that, from (41), the quantity

$$\frac{1}{\mu} \frac{[\phi'(t) - \phi'_{k,\mu}(t/\mu)]^2}{2\phi''(t)} = \frac{\mu}{2\phi''(t)} \left[ \frac{\phi'(t) - \phi'_{k,\mu}(t/\mu)}{\mu} \right]^2 \quad (52)$$



vanishes as  $\mu \rightarrow 0$ , implying that, asymptotically, using (51a) or (51b) in (50) makes no difference. This equivalence is not surprising because, in principle, one can construct an arbitrary number of *distinct* approximations that are *asymptotically* equivalent. The relevant fact is that different forms for the correction correspond to different approximations of the true error probability, which will perform differently for *finite* values of  $1/\mu$ . We shall see in the proof of Theorem 3 why (51b) is a refined version of (51a). From now on, unless otherwise stated, we shall make reference to (51b).

□

REMARK III. Applying (40) and (41), it is immediately seen that:

$$\epsilon_{k,\mu}(\theta_\gamma) = \mathcal{O}(\mu), \quad (53)$$

meaning that the overall correction appearing at the exponent in (50) remains bounded as  $\mu \rightarrow 0$ .

□

REMARK IV. The theorem is stated with reference to large deviations of the type  $\mathbb{P}[\mathbf{y}_{k,\mu}^* > \gamma]$ , with  $\gamma > \mathbb{E}[\mathbf{x}]$ . We notice that the main formulas (50), (51a) and (51b), for the complementary case  $\gamma < \mathbb{E}[\mathbf{x}]$  and  $\mathbb{P}[\mathbf{y}_{k,\mu}^* \leq \gamma]$  remain unchanged. The only difference in the claim is that  $\theta_\gamma < 0$ . These facts can be verified by repeating the proof of the theorem for this complementary case. Alternatively, they can be verified by directly applying the claim of Theorem 3 to the random variable  $-\mathbf{y}_{k,\mu}^*$ , and by observing that the difference between  $<$  and  $\leq$  is immaterial in the proof of the theorem.

□

REMARK V. The proof of Theorem 3 assumes that the local statistic  $\mathbf{x}_k(n)$  is not of lattice type. We recall that the distribution of a lattice random variable with span  $\Delta$  is concentrated at the points  $d, d \pm \Delta, d \pm 2\Delta, \dots$  for a certain real number  $d$  — see, e.g., [40]. The use of the non-lattice assumption in our proof is discussed in the Remark VI at the end of Appendix B. It is useful to observe that, when dealing with the classical and simplest case of normalized sums of i.i.d. random variables, the exact asymptotics for the lattice case must take into account a further correction term that is related to the lattice span [34]. However, as already observed, our asymptotic setting with vanishing step-size is markedly different from that addressed in the literature for the problem of normalized sums of i.i.d. random variables. It is therefore not easy to anticipate if the results of Theorem 3 hold as they are also for the lattice case, if they hold

as they are under additional conditions, or if further corrections are needed.

□

Before ending this section it is useful to collect the steps needed to evaluate (50). Depending on the particular application, the formulas in this listing might need to be evaluated numerically. For practical purposes, the infinite summations appearing in  $\epsilon_{k,\mu}(\cdot)$  must be obviously truncated.

---

**Evaluation of  $\mathcal{P}_{k,\mu}(\gamma)$  in (50).**

---

- 1) Find the solution  $\theta_\gamma$  to the stationary equation:

$$\phi'(\theta_\gamma) = \frac{1}{\theta_\gamma} \sum_{\ell=1}^S \psi(p_\ell \theta_\gamma) = \gamma.$$

- 2) Compute the rate function:

$$\Phi(\gamma) = \sup_{t \in \mathbb{R}} [\gamma t - \phi(t)] = \gamma \theta_\gamma - \phi(\theta_\gamma).$$

- 3) Compute the second derivative:

$$\phi''(\theta_\gamma) = \frac{1}{\theta_\gamma^2} \sum_{\ell=1}^S [(p_\ell \theta_\gamma) \psi'(p_\ell \theta_\gamma) - \psi(p_\ell \theta_\gamma)].$$

- 4) Compute the convergence errors (40) and (41) by using (32) and (38):

$$\begin{aligned} \mathcal{C}_1 &= \phi(\theta_\gamma) - \mu \sum_{i=1}^{\infty} \sum_{\ell=1}^S \psi((1-\mu)^{i-1} b_{k,\ell}(i) \theta_\gamma), \\ \mathcal{C}_2 &= \phi'(\theta_\gamma) - \mu \sum_{i=1}^{\infty} \sum_{\ell=1}^S (1-\mu)^{i-1} b_{k,\ell}(i) \times \\ &\quad \psi'((1-\mu)^{i-1} b_{k,\ell}(i) \theta_\gamma). \end{aligned}$$

- 5) Compute the correction term in (51b):

$$\epsilon_{k,\mu}(\theta_\gamma) = \mathcal{C}_1 + \frac{\mathcal{C}_2^2}{2 \phi''(\theta_\gamma)}.$$

- 6) Using the above quantities, evaluate  $\mathcal{P}_{k,\mu}(\gamma)$ .
-

#### IV. IMPORTANCE SAMPLING AND CRAMÉR'S TRANSFORM

Real-world detectors are usually required to be high-performing, i.e., they must exhibit very low error probabilities. Unfortunately, standard Monte Carlo techniques to estimate their performance become unfeasible for error probability values in the order of  $10^{-6}$ . To overcome this issue, we shall resort to *importance sampling* techniques [42], which can dramatically reduce the number of runs needed to reach a prescribed level of estimation accuracy. We start by illustrating briefly the importance sampling philosophy. Then, in the next section, we show how it should be applied to the adaptive distributed detection problem.

##### A. Cramér's Transform

Let us refer to a random variable  $\mathbf{y}$ , assumed continuous for ease of description. We shall denote by  $f(y)$  its probability density function (pdf). Consider now another pdf  $\tilde{f}(y)$  that does not vanish (except for zero-measure sets) when  $f(y) > 0$ , and introduce the following weighting function:

$$w(y) = \frac{f(y)}{\tilde{f}(y)}, \quad (54)$$

that is, the likelihood ratio between  $f(y)$  and  $\tilde{f}(y)$ . It then holds that:

$$\begin{aligned} \mathbb{P}[\mathbf{y} > \gamma] &= \int_{\gamma}^{\infty} f(y) dy = \int_{\gamma}^{\infty} w(y) \tilde{f}(y) dy \\ &= \mathbb{E}_{\tilde{f}} [w(\mathbf{y}) \mathcal{I}_{\{\mathbf{y} > \gamma\}}], \end{aligned} \quad (55)$$

where  $\mathcal{I}_{\mathcal{E}}$  is the indicator of an event  $\mathcal{E}$ , and  $\mathbb{E}_{\tilde{f}}[\cdot]$  denotes expectation computed over the *transformed* pdf  $\tilde{f}(y)$ . The above equation shows that the quantity to be estimated can be regarded as the expectation, under the transformed pdf, of the indicator of the event  $\{\mathbf{y} > \gamma\}$ , *weighted* by the function  $w(y)$ . The rationale behind importance sampling is that, by an appropriate choice of the weighting function, it is possible to map an event that is rare under the original sampling pdf  $f(y)$ , into an event that is *not* rare under the new sampling pdf  $\tilde{f}(y)$ . In this way, the number of Monte Carlo iterations needed to estimate the expectation is reduced, because (important) samples are generated around the body (not the tail) of the new distribution. An accurate estimate of the probability tails is enabled by the weighting function  $w(y)$ .

When working with random variables obeying a LDP, there is a classical way to select the transformed pdf  $\tilde{f}(y)$ . This is usually referred to as *exponential twisting* of  $f(y)$ , and amounts

to selecting [34], [35]:

$$\tilde{f}(y) = e^{\eta y - \ln \mathbb{E}[e^{\eta \mathbf{y}}]} f(y) \Leftrightarrow w(y) = e^{-\eta y + \ln \mathbb{E}[e^{\eta \mathbf{y}}]}. \quad (56)$$

When  $\mathbf{y}$  is not a continuous random variable, the exponential twisting can be rephrased in terms of probability measures  $m$  and  $\tilde{m}$  as:

$$\tilde{m}(dy) = e^{\eta y - \ln \mathbb{E}[e^{\eta \mathbf{y}}]} m(dy). \quad (57)$$

The results presented in this section hold, *mutatis mutandis*, for this setting.

The aforementioned change of measure was originally proposed by Cramér [41] to compute large deviations exponents, and is accordingly also known as Cramér’s transform. The choice of a given exponential twisting (i.e., the choice of the parameter  $\eta$ ) is critical in determining the accuracy of the estimates produced by the importance sampling algorithm. Interestingly, theoretical studies suggest to use for importance sampling in general exactly the *same* exponential twisting needed to compute the error exponents — see, e.g., [42]. For the classical and simplest case of summation of i.i.d. random variables, the latter kind of exponential twisting is well-known — see, e.g., [34], [35]. For our specific problem of adaptive distributed detection, the solution is more involved, and will be discussed in the next section. In particular, it will be shown that the theoretical results established in this paper are crucial to enable an accurate design of the importance sampling simulations.

### B. Importance Sampling for Adaptive Distributed Detection

Since we are interested in evaluating the steady-state performance of the adaptive distributed algorithm, it is sufficient to examine the behavior, for  $n$  sufficiently large, of the random variable  $\mathbf{y}_k^*(n)$  introduced in (21), which, according to (22), *has the same distribution of* the diffusion output with null transient.

In order to implement the importance sampling recipe, we need to perform the exponential twisting (56). As already mentioned, the parameter  $\eta$  will be chosen as that value corresponding to the change of measure used to compute the error exponents. For the specific case of adaptive distributed detection, this exponential twisting is extensively discussed and employed in Appendix B — see, e.g., (143) and (178), and amounts to selecting  $\eta = \theta_\gamma / \mu$ , where  $\theta_\gamma$  is the solution to

the stationary equation (48) in Theorem 3. Denoting by

$$\phi_{k,\mu}(t; n) \triangleq \ln \mathbb{E}[e^{t\mathbf{y}_k^*(n)}] = \sum_{i=1}^n \sum_{\ell=1}^S \psi(\mu(1-\mu)^{i-1}b_{k,\ell}(i)t), \quad (58)$$

the LMGF of the random variable  $\mathbf{y}_k^*(n)$  (the additive form above comes simply from the i.i.d. assumption on the  $\{\mathbf{x}_\ell(i)\}$ ), the choice  $\eta = \theta_\gamma/\mu$  applied in (56) then yields:

$$w(y) = e^{-\frac{\theta_\gamma}{\mu} y + \phi_{k,\mu}(\frac{\theta_\gamma}{\mu}; n)}. \quad (59)$$

Unfortunately, in order to generate samples according to the transformed pdf  $\tilde{f}(y)$  in (56), the weighting function  $w(y)$  is not sufficient, since one needs to know also the pdf  $f(y)$  of the random variable  $\mathbf{y}_k^*(n)$ . As already observed, knowledge of this distribution is seldom available. To overcome this issue, one may consider running first a Monte Carlo simulation by generating random instances of the local statistics  $\mathbf{x}_\ell(i)$ , and then evaluating  $\mathbf{y}_k^*(n)$  through (21). However, recall that we are interested in implementing importance sampling. This means that we must find the appropriate change of measure applied to  $\mathbf{x}_\ell(i)$  (and not to  $\mathbf{y}_k^*(n)$ ) that would produce the desired exponential change of measure (59) on the random variable  $\mathbf{y}_k^*(n)$ . To this aim, let us denote by  $p(x)$  the pdf of  $\mathbf{x}_\ell(i)$ . We now show that the desired goal can be achieved by drawing the  $(i, \ell)$ -th sample  $\mathbf{x}_\ell(i)$  from the pdf

$$\tilde{p}_{i,\ell}(x) = e^{\eta_{i,\ell}x - \psi(\eta_{i,\ell})}p(x), \quad (60)$$

with

$$\eta_{i,\ell} \triangleq (1-\mu)^{i-1}b_{k,\ell}(i)\theta_\gamma. \quad (61)$$

All these samples are still generated independently, but now they are no longer identically distributed. Note also that the above transformation depends upon the index  $k$  of the agent under consideration, even if the subscript has been suppressed for ease of notation.

To see why the choice (60) will correspond to (59), let us introduce the joint ensemble  $\mathbf{X} = \{\mathbf{x}_\ell(i)\}$ , for  $i = 1, 2, \dots, n$  and for  $\ell = 1, 2, \dots, S$ . The expectation in (55) can then be rewritten in terms of the distribution of  $\mathbf{X}$  as follows:

$$\mathbb{P}[\mathbf{y}_k^*(n) > \gamma] = \mathbb{E}[\mathcal{I}_{\{\mathbf{y}_k^*(n) > \gamma\}}] = \mathbb{E}_{\tilde{p}} \left[ \frac{p(\mathbf{X})}{\tilde{p}(\mathbf{X})} \mathcal{I}_{\{\mathbf{y}_k^*(n) > \gamma\}} \right]. \quad (62)$$

But since the  $\mathbf{x}_\ell(i)$  are spatially and temporally i.i.d., we have from (60) and (59):

$$\begin{aligned} \frac{p(\mathbf{X})}{\tilde{p}(\mathbf{X})} &= e^{-\sum_{i=1}^n \sum_{\ell=1}^S \eta_{i,\ell} \mathbf{x}_\ell(i) + \sum_{i=1}^n \sum_{\ell=1}^S \psi(\eta_{i,\ell})} \\ &= e^{-\frac{\theta_\gamma}{\mu} \mathbf{y}_k^*(n) + \phi_{k,\mu}\left(\frac{\theta_\gamma}{\mu}; n\right)} = w(\mathbf{y}_k^*(n)), \end{aligned} \quad (63)$$

where, in the last equality, we applied definitions (21) and (58). This result shows that (62) corresponds to the expectation in (55) with the pdf  $\tilde{f}(y)$  chosen as in (59).

Before ending this section, we stress that the above considerations show one further important benefit: *our large deviations results about distributed detection over adaptive networks are also very useful in performing a careful design of a Monte Carlo simulator based on importance sampling.*

## V. ILLUSTRATIVE DESIGN

### A. Network Topology and Combination Weights

We consider a network made of  $S = 10$  sensors, arranged so as to form the topology in the inset of Fig. 1. Given the topology, two different combination matrices will be tested. The first one is defined by the so-called Metropolis rule [7]. Denoting by  $\mathcal{N}_k$  the neighborhood of the  $k$ -th agent (including  $k$  itself), and by  $n_k$  the cardinality  $|\mathcal{N}_k|$  (aka the degree of the  $k$ -th agent), the Metropolis rule is defined by:

$$a_{k,\ell} = \begin{cases} 1/\max\{n_k, n_\ell\}, & \ell \in \mathcal{N}_k \setminus \{k\}, \\ 1 - \sum_{m \in \mathcal{N}_k \setminus \{k\}} a_{k,m}, & \ell = k, \\ 0, & \ell \notin \mathcal{N}_k. \end{cases} \quad (64)$$

This choice provides a doubly-stochastic  $A$ , and, hence, the corresponding Perron eigenvector has uniform entries,  $p_\ell = 1/S$  for all  $\ell = 1, 2, \dots, S$ .

The second combination policy is the uniform averaging rule [7]:

$$a_{k,\ell} = \begin{cases} 1/n_k, & \ell \in \mathcal{N}_k, \\ 0, & \ell \notin \mathcal{N}_k. \end{cases} \quad (65)$$

This choice provides a right-stochastic  $A$ , whose Perron eigenvector is available in closed form, and has entries given by [7]:

$$p_\ell = \frac{n_\ell}{\sum_{m=1}^S n_m}, \quad \ell = 1, 2, \dots, S. \quad (66)$$

### B. Hypothesis Test, Local Statistics and Simulation

We examine the following canonical shift-in-mean detection problem with noise distributed according to a Laplace distribution, which is considered in [1], [2] as well. The Laplace pdf (with scale parameter set to 1, without loss of generality) will be denoted by:

$$\mathcal{L}(d) = \frac{1}{2}e^{-|d|}. \quad (67)$$

The hypothesis test can then be formulated as follows:

$$\mathcal{H}_0 : \mathbf{d}_k(n) \sim \mathcal{L}(d), \quad (68)$$

$$\mathcal{H}_1 : \mathbf{d}_k(n) \sim \mathcal{L}(d - \rho), \quad (69)$$

where  $\mathbf{d}_k(n)$  denotes the measurement collected by agent  $k$  at time  $n$ , and  $\rho > 0$  is the shift-in-mean parameter. The local statistics  $\mathbf{x}_k(n)$  are chosen as the local log-likelihood ratios:

$$\mathbf{x}_k(n) = \ln \left( \frac{\mathcal{L}(\mathbf{d}_k(n) - \rho)}{\mathcal{L}(\mathbf{d}_k(n))} \right) = |\mathbf{d}_k(n)| - |\mathbf{d}_k(n) - \rho|. \quad (70)$$

Even if the above relationship shows clearly how to obtain the random variable  $\mathbf{x}_k(n)$  from the knowledge of  $\mathbf{d}_k(n)$ , it is useful, for later use, to evaluate explicitly the distribution of  $\mathbf{x}_k(n)$  under the two hypotheses. To this aim, we rewrite (70) as:

$$\mathbf{x}_k(n) = \begin{cases} -\rho, & \mathbf{d}_k(n) < 0, \\ +\rho, & \mathbf{d}_k(n) > \rho, \\ 2\mathbf{d}_k(n) - \rho, & \mathbf{d}_k(n) \in [0, \rho], \end{cases} \quad (71)$$

which is obtained by using, in the three ranges considered in (71), the explicit definition of the absolute values appearing in (70). We see then that  $\mathbf{x}_k(n)$  is a random variable of mixed type, taking values in the range  $[-\rho, \rho]$ , and with two atoms located at  $\pm\rho$ . For ease of description, we find convenient to use, for random variables of mixed type, the generalized pdf written using the Dirac-delta function  $\delta(x)$ .

Accordingly, let  $p_0(x)$  and  $p_1(x)$  denote the generalized pdfs of  $\mathbf{x}_k(n)$  under  $\mathcal{H}_0$  and  $\mathcal{H}_1$ , respectively. For a shift-in-mean with respect to a symmetric pdf (as  $\mathcal{L}(d)$  is), it is well-known

that the log-likelihood ratio exhibits the symmetry property:  $p_1(x) = p_0(-x)$  — see, e.g., [44]. Thus, it suffices to focus on  $p_0(x)$ . To this aim, we observe that, in view of (71), the cumulative distribution function of  $\mathbf{x}_k(n)$  is

$$\mathbb{P}_0[\mathbf{x}_k(n) \leq x] = \begin{cases} 0, & x < -\rho, \\ \mathbb{P}_0\left[\mathbf{d}_k(n) \leq \frac{x+\rho}{2}\right], & x \in [-\rho, \rho), \\ 1, & x \geq \rho, \end{cases} \quad (72)$$

which corresponds to the following generalized pdf:

$$\begin{aligned} p_0(x) &= \mathbb{P}_0[\mathbf{d}_k(n) < 0]\delta(x + \rho) + \mathbb{P}_0[\mathbf{d}_k(n) > \rho]\delta(x - \rho) + \\ &\quad \frac{1}{2}\mathcal{L}\left(\frac{x+\rho}{2}\right)\Pi\left(\frac{x}{2\rho}\right) \\ &= \frac{1}{2}\delta(x + \rho) + \frac{e^{-\rho}}{2}\delta(x - \rho) + \frac{1}{4}e^{-\frac{x+\rho}{2}}\Pi\left(\frac{x}{2\rho}\right), \end{aligned} \quad (73)$$

where  $\Pi(x)$  is a unit-width rectangular window centered at 0. The corresponding LMGF of  $\mathbf{x}_k(n)$  is computable in closed form:

$$\begin{aligned} \psi_0(t) &= \ln \mathbb{E}_0[e^{t\mathbf{x}_k(n)}] = \ln \left( \int_{-\infty}^{\infty} e^{tx} p_0(x) dx \right) \\ &= \ln \left( \frac{e^{-t\rho}}{2} + \frac{e^{(t-1)\rho}}{2} + \frac{e^{-\rho/2}\rho}{2} \operatorname{sinh}[\rho(t - 1/2)] \right), \end{aligned} \quad (74)$$

where

$$\operatorname{sinh}(x) \triangleq \frac{\sinh(x)}{x}, \quad \operatorname{sinh}(0) = 1. \quad (75)$$

Since, as observed,  $p_1(x) = p_0(-x)$ , the LMGF under  $\mathcal{H}_1$  is easily obtained as  $\psi_1(t) = \psi_0(-t)$ . We note in passing that the above explanation and the following derivations can be restated in a more formal language by using, for the mixed-type random variable  $\mathbf{x}_k(n)$ , a probability measure made of the superposition of two singular, atomic measures (with masses located at  $\pm\rho$ ), and an absolutely continuous measure with density given by the third term in (73).

Before concluding this section, it remains to show how to implement, for the considered Laplace example, the importance sampling method described in Sec. IV. We shall focus on



hypothesis  $\mathcal{H}_0$ , and, again, the results for  $\mathcal{H}_1$  can be simply obtained from the relationship  $p_1(x) = p_0(-x)$ . Applying an exponential twisting with parameter  $\eta$  to the generalized pdf in (73), we get:

$$\begin{aligned}\tilde{p}_0(x) &= e^{\eta x - \psi_0(\eta)} p_0(x) \\ &= \frac{e^{-\eta\rho}}{2e^{\psi_0(\eta)}} \delta(x + \rho) + \\ &\quad \frac{e^{(\eta-1)\rho}}{2e^{\psi_0(\eta)}} \delta(x - \rho) + \\ &\quad \frac{e^{-\rho/2}}{4e^{\psi_0(\eta)}} e^{x(\eta-1/2)} \Pi\left(\frac{x}{2\rho}\right).\end{aligned}\tag{76}$$

Introducing the definitions:

$$p_- = \frac{e^{-\eta\rho}}{2e^{\psi_0(\eta)}}, \quad p_+ = \frac{e^{(\eta-1)\rho}}{2e^{\psi_0(\eta)}},\tag{77}$$

and using (74), by simple algebra relation (76) becomes:

$$\begin{aligned}\tilde{p}_0(x) &= p_- \delta(x + \rho) + p_+ \delta(x - \rho) + [1 - p_- - p_+] \times \\ &\quad \frac{1}{2\rho} \frac{e^{x(\eta-1/2)}}{\sinh[\rho(\eta-1/2)]} \Pi\left(\frac{x}{2\rho}\right).\end{aligned}\tag{78}$$

Generating a random variable  $\mathbf{x}_k(n)$  according to the distribution  $\tilde{p}_0(x)$  is now an easy task. As a matter of fact, (78) reveals that  $\tilde{p}_0(x)$  is a mixture of three components such that  $\mathbf{x}_k(n)$  takes on the value  $-\rho$  with probability  $p_-$ , the value  $+\rho$  with probability  $p_+$ , and that otherwise it must be sampled from the pdf

$$\frac{1}{2\rho} \frac{e^{x(\eta-1/2)}}{\sinh[\rho(\eta-1/2)]} \Pi\left(\frac{x}{2\rho}\right).\tag{79}$$

### C. A Normal Approximation

In the following analysis, we shall compare the error probabilities estimated empirically, to the refined asymptotic formulas obtained in the present work. As a further term of comparison, we would like to add a normal approximation that will follow from the asymptotic normality result proved in [1], [2]. Actually, this result was obtained there for doubly-stochastic connection matrices, but the generalization to the case of right-stochastic matrices comes essentially at no cost and can be stated as:

$$\boxed{\frac{\mathbf{y}_{k,\mu}^* - \mathbb{E}[\mathbf{x}]}{\sqrt{\mu \sigma_{\text{lim}}^2}} \xrightarrow[\rightsquigarrow]{\mu \rightarrow 0} \mathcal{N}(0, 1)}\tag{80}$$

where the limiting variance  $\sigma_{\text{lim}}^2$  can be obtained by applying result (92) to the second cumulant (i.e., the variance) of  $\mathbf{y}_{k,\mu}^*$ , which yields:

$$\frac{\text{VAR}[\mathbf{y}_{k,\mu}^*]}{\mu} \xrightarrow{\mu \rightarrow 0} \frac{\sigma_x^2}{2} \sum_{\ell=1}^S p_\ell^2 \triangleq \sigma_{\text{lim}}^2. \quad (81)$$

We emphasize that this result is consistent with previous findings obtained in the context of mean-square-error estimation — see, e.g., [6].

We see from (81) that the ratio  $\text{VAR}[\mathbf{y}_{k,\mu}^*]/(\mu\sigma_{\text{lim}}^2)$  converges to one as  $\mu$  goes to zero. In view of Slutsky's Theorem [47], this fact implies that the following alternative version of the convergence in distribution in (80) holds:

$$\boxed{\frac{\mathbf{y}_{k,\mu}^* - \mathbb{E}[\mathbf{x}]}{\sqrt{\text{VAR}[\mathbf{y}_{k,\mu}^*]} } \xrightarrow[\sim]{\mu \rightarrow 0} \mathcal{N}(0, 1)} \quad (82)$$

While the two formulations (80) and (82) are asymptotically equivalent, it is expected that (82) offers a better performance since it replaces the asymptotic variance with the actual one. The relationship shown in (82) suggests the following way to approximate the probability  $\mathbb{P}[\mathbf{y}_{k,\mu}^* > \gamma]$ :

$$\mathbb{P}[\mathbf{y}_{k,\mu}^* > \gamma] \approx Q\left(\frac{\gamma - \mathbb{E}[\mathbf{x}]}{\sqrt{\text{VAR}[\mathbf{y}_{k,\mu}^*]}}\right), \quad (83)$$

where  $Q(\cdot)$  denotes the complementary cumulative distribution function of a standard normal distribution.

#### D. Analysis of the Results

Let us now examine the adaptive distributed network of detectors in operation, by reporting the evidence arising from our Monte Carlo analysis. We refer to a sufficiently large time horizon, such that the steady-state assumption applies, and evaluate the error probabilities for different values of the step-size.

The exact asymptotics provided by Theorem 3 will be computed by implementing the six-steps recipe described at the end of Sec. III. Clearly, in doing so, we must use the LMGF  $\psi_0(t)$  if we are working under  $\mathcal{H}_0$ , and the LMGF  $\psi_1(t)$  if we are working under  $\mathcal{H}_1$ . The normal approximation will be instead obtained as described in the previous section — see (83).

In the two examples that we are going to discuss we choose the detection threshold as detailed in [1], [2], obtaining  $\gamma = 0$ . This implies, by the symmetry property  $p_1(x) = p_0(-x)$ , that the

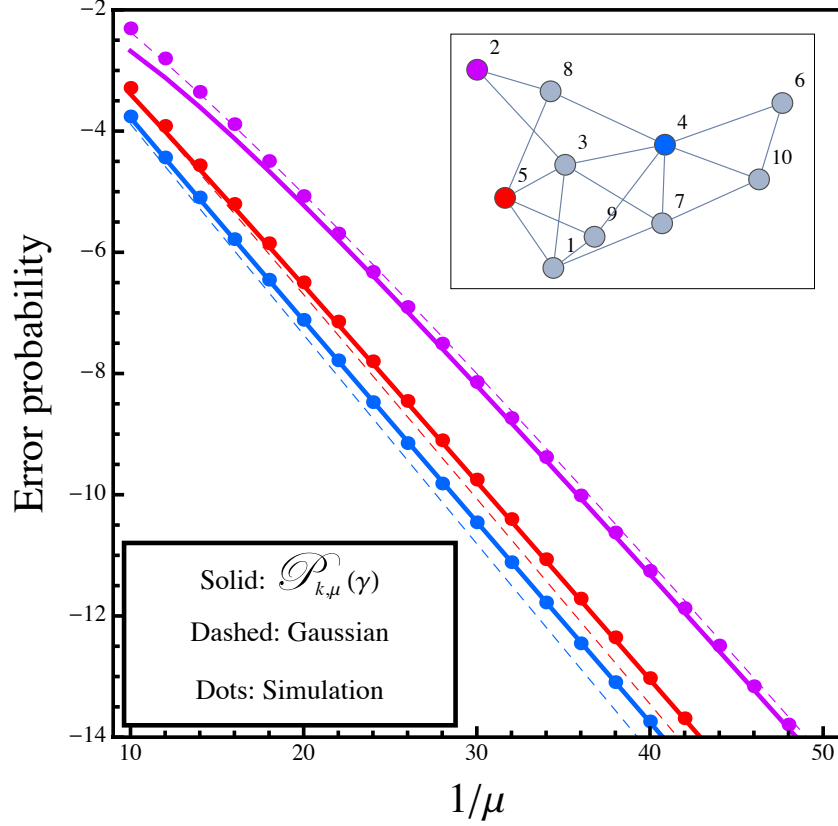


Fig. 1. Laplace example discussed in Sec. V-B, with  $\rho = 0.6$ . The network topology is depicted in the inset plot, and the combination weights  $a_{k,\ell}$  follow the Metropolis rule (64). The performance of agents 2, 4 and 5 is displayed. Dots refer to the empirical steady-state error probabilities  $\alpha_{k,\mu} = \beta_{k,\mu}$  at different sensors, obtained via Monte Carlo simulation with importance sampling, as described in IV. Solid curves refer to the exact asymptotics provided by Theorem 3. Dashed curves refer to the normal approximation (83).

error probabilities of first and second kind defined by (28) are equal, namely:

$$\alpha_{k,\mu} = \beta_{k,\mu}. \quad (84)$$

Consistently, in the following description, the terminologies “error probability” and “error exponent” refer to any of these errors.

We start by considering the doubly-stochastic combination matrix obtained with the Metropolis rule (64). In Fig. 1, the performance of the agents is displayed as a function of  $1/\mu$ , and different agents are marked with different colors. The main system features, which were already commented in [1], [2], are here summarized.

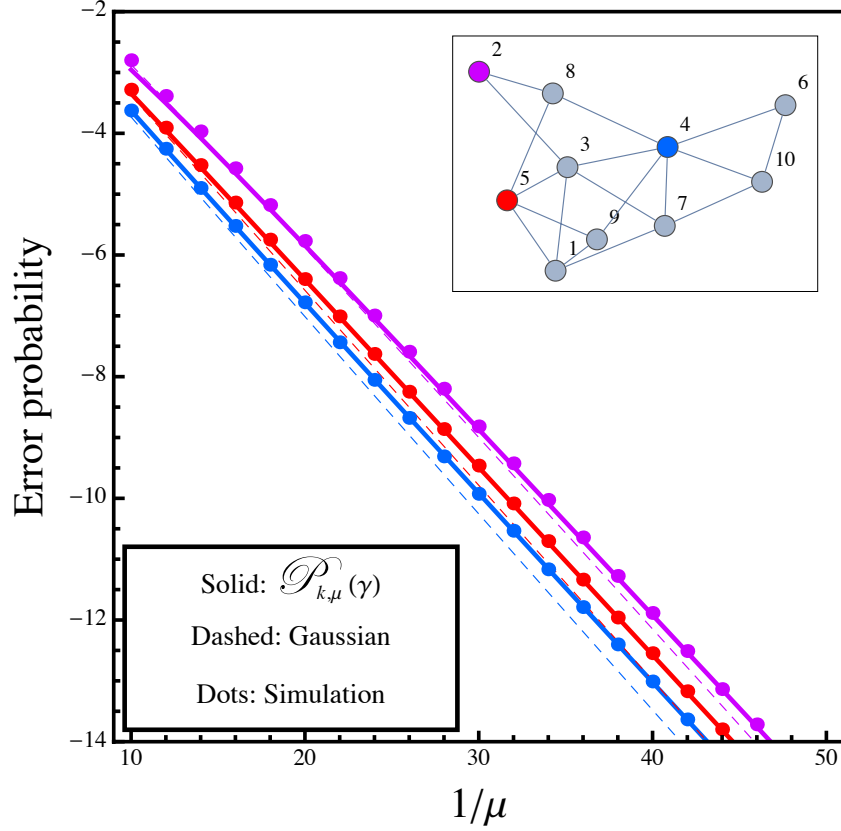


Fig. 2. Laplace example discussed in Sec. V-B, with  $\rho = 0.6$ . The network topology is depicted in the inset plot, and the combination weights  $a_{k,\ell}$  follow the averaging rule (65). The performance of agents 2, 4 and 5 is displayed. Dots refer to the empirical steady-state error probabilities  $\alpha_{k,\mu} = \beta_{k,\mu}$  at different sensors, obtained via Monte Carlo simulation with importance sampling, as described in IV. Solid curves refer to the exact asymptotics provided by Theorem 3. Dashed curves refer to the normal approximation (83).

The first evidence is that the different curves pertaining to different agents stay nearly parallel for sufficiently small values of the step-size  $\mu$ , i.e., the detection error probabilities at different sensors vanish exponentially as functions of  $1/\mu$ , sharing the same detection error exponent. This was the main result revealed by the large deviations analysis performed in [1], [2].

However, as already noticed in [1], [2], the large deviations tool is not powerful enough to capture an important feature of the distributed behavior. Indeed, the second evidence emerging from the simulations is that the error probability curves in Fig. 1 are basically ordered, and the ordering is closely related to the network connection structure. Comparing the detection performance of three specific agents, namely, agents 2, 4, 5, we see that the ordering reflects the

degree of connectivity of each agent. For instance, agent 4 has the highest number of neighbors, and its performance is the best one, while agent 2 is the most isolated, and its error probability curve appears consistently as the highest one. According to what one expects, agent 5 is in an intermediate position. *The new fact here is that, using the results of the current manuscript, we are now able to provide a systematic analysis of the above features, as well as of the exact interplay with network connectivity.*

First of all, what in [1], [2] was only a partial evidence arising from a particular numerical experiment, emerges now, thanks to the refined asymptotic analysis, as the *universal behavior of adaptive distributed detection*.

Moreover, the refined asymptotic approximations provided by Theorem 3 can be used to obtain quantitative predictions of the actual system performance, as we proceed to explain. The refined formulas are represented by the solid curves in Fig. 1. We see that the empirical probability points (the dots) converge toward the theoretical solid curves as the step-size  $\mu$  decreases (i.e., as we move to the right in the plot). Remarkably, the theoretical formulas provided by our theorems are able to embody the dependencies between the network connection structure and the detection performance at different sensors, as it is witnessed by the correct ordering of the curves. In addition, a gap between the exact asymptotics (solid curves) and the normal approximation (dashed curves) is clearly observed. This should come as no surprise, since the normal approximation is expected to be accurate when working with *small* deviations, and must accordingly provide a wrong prediction in the large deviations regime — see also the discussion in [1], [2].

We now switch to the analysis of the right-stochastic combination matrix provided by the uniform averaging rule (65). The corresponding results are reported in Fig. 2. First, we note that, as in the doubly-stochastic case, the error probability curves vanish exponentially fast as functions of  $1/\mu$ , and stay nearly parallel as  $\mu$  goes to zero. This is consistent with the prediction that they share the same error exponent, as dictated by Theorem 2 for the general case of *right-stochastic* combination matrices. Also in this example, we are able to appreciate the goodness of the refined approximations obtained with Theorem 3, and the fact that the empirical points depart from the normal approximation.

To get further insights, we now compare the detection performance of the two aforementioned

combination matrices. To begin with, we compute the error exponents pertaining to the two systems. For the particular example considered, we obtain (recall that the threshold is set to  $\gamma = 0$ ):

$$\Phi^{\text{DS}}(0) \approx 0.75 > \Phi^{\text{RS}}(0) \approx 0.7, \quad (85)$$

for the doubly-stochastic and the right-stochastic case, respectively. These values appear to suggest that the doubly-stochastic combination policy asymptotically outperforms the right-stochastic combination policy. This conclusion may somehow be expected because, asymptotically, a doubly-stochastic combination policy weights the local statistics equally, while a right-stochastic combination policy does not. In the presence of i.i.d. observations, the former strategy seems to be preferable. To see if this first-order analysis suffices, let us now apply the refined asymptotic formulas.

Accordingly, in Fig. 3 we display the theoretical error probability curves obtained with Theorem 3, for the Metropolis combination matrix (doubly-stochastic case, line and markers) and for the uniform averaging combination matrix (right-stochastic case, solid curves). An interesting behavior arises. *We observe that the relative performance of the two different combination policies depends strongly on the connectivity of the individual agent.* For instance, for the well-connected agent 4, the doubly-stochastic combination policy delivers superior performance, while exactly the converse is true for the scarcely connected agent 2. Moreover, if we consider as network performance the arithmetic average of the error probabilities (black curves), we see that the right-stochastic combination policy is globally superior.

An explanation for this behavior is as follows. Denoting by  $\lambda_2^{\text{DS}}$  and  $\lambda_2^{\text{RS}}$  the second largest magnitude eigenvalues of the doubly-stochastic and right-stochastic combination matrices, respectively, we see that

$$\lambda_2^{\text{DS}} \approx 0.83 > \lambda_2^{\text{RS}} \approx 0.7, \quad (86)$$

implying that the right-stochastic weights will converge to the corresponding Perron eigenvector faster than the doubly-stochastic weights. Examining the detection performance, we see from here that the slower convergence of the doubly-stochastic combination matrix has a detrimental effect on the less connected agents. We could say that the benefits of the higher (doubly-stochastic case) exponent are more than compensated by the faster (right-stochastic case) convergence to the steady-state behavior.

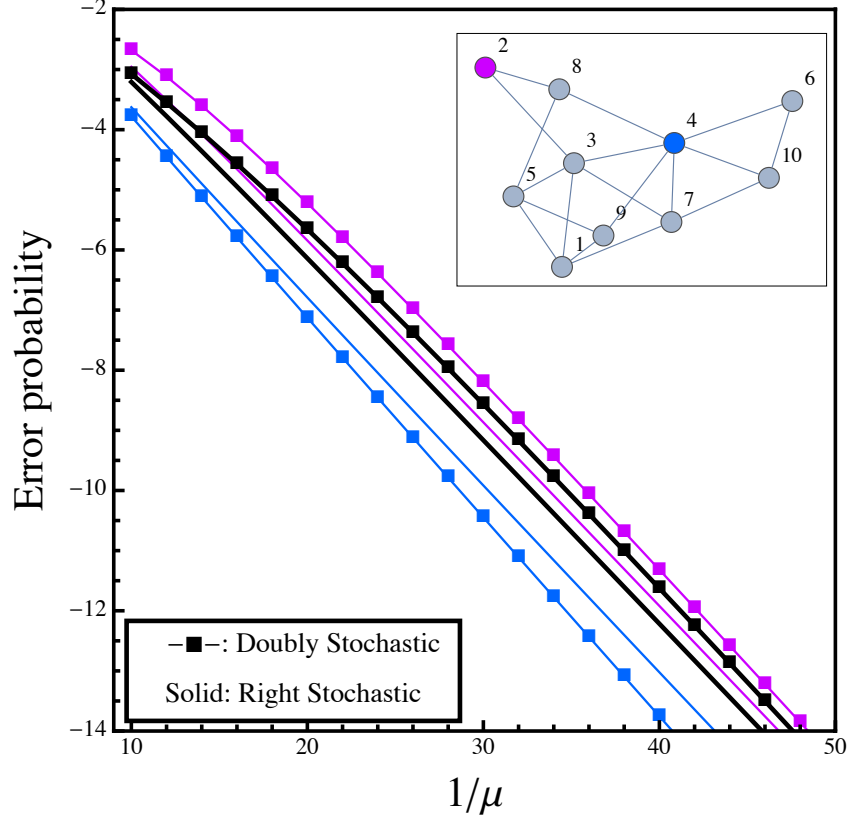


Fig. 3. Laplace example discussed in Sec. V-B, with  $\rho = 0.6$ . The network topology is depicted in the inset plot, and the performance corresponding to two combination matrices is compared, namely, the doubly-stochastic matrix obtained with the Metropolis rule (64), and the right-stochastic matrix obtained with the uniform averaging rule (65). The performance of agents 2 and 4 is displayed, along with the average network performance, namely, the arithmetic mean of the error probabilities of all agents (black). All curves are computed by using the exact asymptotics provided by Theorem 3.

Notice that the observed behavior is not in contrast with what is predicted by (85). Based on an analysis at the first leading order in the exponent, the doubly-stochastic combination policy are asymptotically superior to the right-stochastic one. This means that the two curves corresponding to sensor 2 in the figure must cross for a certain vanishingly small  $\mu$ . What the refined analysis is able to tell is that this value might be too small for a given regime of analysis.

Before concluding, we would like to stress that the above comparison between combination policies should be considered preliminary, in that: *i*) the comparison has been made for two systems operating with the same value of the step-size  $\mu$ , and, for a more complete view, the analysis should be complemented by examining also the transient behavior of the two

combination policies; *ii*) the evidences are obtained with reference to a particular doubly-stochastic matrix and a particular right-stochastic matrix. This notwithstanding, while no general conclusions can be drawn at this stage regarding the relative advantages of the two kinds of combination strategies for individual agents, a general trend seems to emerge: *connectivity matters*. Since the connectivity features are embodied in the higher-order corrections, the simplest large deviations analysis is not sufficient, and the refined exact asymptotics provided by Theorem 3 are crucial in assessing the performance of adaptive distributed detection over networks.

## APPENDIX A

In the following, the  $r$ -th derivative of a function  $f(t)$  will be denoted by  $f^{(r)}(t)$ , with the convention that  $f^{(0)}(t) = f(t)$ . When convenient, the first three derivatives will be alternatively denoted by  $f'(t)$ ,  $f''(t)$ , and  $f'''(t)$ . Moreover, the notation  $f_\mu = \mathcal{O}(\mu)$  means that the ratio  $f_\mu/\mu$  stays bounded as  $\mu \rightarrow 0$ .

**THEOREM 1 (Fundamental properties of  $\phi_{k,\mu}^{(r)}(t)$ ).** *Assume that  $\psi(t) < \infty$  for all  $t \in \mathbb{R}$ , and introduce the quantity:*

$$\xi_{i,\ell} \triangleq \mu(1 - \mu)^{i-1} b_{k,\ell}(i). \quad (87)$$

*Then, the following facts hold:*

*i) The LMGF of  $\mathbf{y}_{k,\mu}^*$  can be computed as:*

$$\phi_{k,\mu}(t) = \sum_{i=1}^{\infty} \sum_{\ell=1}^S \psi(\xi_{i,\ell} t) \quad (88)$$

*and the limiting LMGF  $\phi(t)$  is:*

$$\phi(t) = \sum_{\ell=1}^S \omega(p_\ell t) \quad (89)$$

*ii) For  $r = 1, 2, \dots$ , it holds that:*

$$\phi_{k,\mu}^{(r)}(t) = \sum_{i=1}^{\infty} \sum_{\ell=1}^S \xi_{i,\ell}^r \psi^{(r)}(\xi_{i,\ell} t) \quad (90)$$

*and*

$$\lim_{\mu \rightarrow 0} \frac{\phi_{k,\mu}^{(r)}(t/\mu)}{\mu^{r-1}} = \phi^{(r)}(t) \quad (91)$$



In particular, the  $r$ -th cumulant,  $\phi_{k,\mu}^{(r)}(0)$ , of the steady-state random variable  $\mathbf{y}_{k,\mu}^*$  satisfies:

$$\lim_{\mu \rightarrow 0} \frac{\phi_{k,\mu}^{(r)}(0)}{\mu^{r-1}} = \phi^{(r)}(0) = \frac{\psi^{(r)}(0)}{r} \sum_{\ell=1}^S p_{\ell}^r \quad (92)$$

where  $\psi^{(r)}(0)$  is the  $r$ -th cumulant of the local statistic  $\mathbf{x}_k(n)$ .

iii) With reference to the convergence in (89) and (91), the following refined estimate of the convergence error holds for  $r = 0, 1, \dots$ :

$$\frac{\phi_{k,\mu}^{(r)}(t/\mu)}{\mu^{r-1}} = \phi^{(r)}(t) + \mathcal{O}(\mu). \quad (93)$$

□

In order to establish the validity of Theorem 1, we start by proving a couple of useful lemmas. Our first lemma is a simple generalization of Lemmas 1 and 2 from [1], [2]. We shall focus on a function  $f(t)$  twice differentiable in  $\mathbb{R}$ , with  $f(0) = 0$ . For such a function, we have that:

- The function  $f(t)/t$  is continuous for all  $t \in \mathbb{R}$ . For any  $t \neq 0$ , the result follows from the continuity of  $f(t)$ . For  $t = 0$ , the result is easily verified by recalling that  $f(0) = 0$ , yielding:

$$\lim_{t \rightarrow 0} \frac{f(t)}{t} = f'(0). \quad (94)$$

- The derivative  $(d/dt)(f(t)/t)$  is continuous for all  $t \in \mathbb{R}$ . For any  $t \neq 0$ , this result follows immediately from the assumed smoothness properties of  $f(t)$ . For  $t = 0$ , the result is easily verified by recalling that  $f(0) = 0$ , and observing that

$$\lim_{t \rightarrow 0} \frac{d}{dt} \frac{f(t)}{t} = \lim_{t \rightarrow 0} \frac{f'(t)t - f(t)}{t^2} = \frac{f''(0)}{2}, \quad (95)$$

where we used L'Hospital's rule [43].

We introduce the auxiliary functions:

$$h_1(t) = \frac{t^2}{2} \times \begin{cases} \max_{\tau \in [0, t]} \left| \frac{d}{d\tau} \frac{f(\tau)}{\tau} \right|, & t \geq 0, \\ \max_{\tau \in [t, 0]} \left| \frac{d}{d\tau} \frac{f(\tau)}{\tau} \right|, & t < 0. \end{cases} \quad (96)$$

and

$$h_2(t) = |t| \times \begin{cases} \max_{\tau \in [0, t]} |f'(\tau)|, & t \geq 0, \\ \max_{\tau \in [t, 0]} |f'(\tau)|, & t < 0. \end{cases} \quad (97)$$

We can easily show that:

$$0 \leq h_1(t) < \infty, \quad 0 \leq h_2(t) < \infty, \quad \forall t \in \mathbb{R}. \quad (98)$$

Indeed,  $h_1(t) \geq 0$  and  $h_2(t) \geq 0$  by definition. Finiteness of both functions follows from Weierstrass extreme value theorem [43] since, by the properties for  $f(t)$  and  $f(t)/t$  discussed above, the maxima appearing in (96) and (97) are maxima of continuous functions over compact sets for any finite  $t$ .

**LEMMA 1.** *Let  $f(t)$  be twice continuously differentiable in  $\mathbb{R}$ , with  $f(0) = 0$ . Let, for all  $i \in \mathbb{N}$ :*

$$0 \leq b_i \leq 1, \quad b_i \xrightarrow{i \rightarrow \infty} p > 0, \quad \text{with } |b_i - p| \leq \mathcal{C}\lambda^i \quad (99)$$

for some  $\mathcal{C} > 0$  and  $0 < \lambda < 1$ . Then, we have, for all  $t \in \mathbb{R}$ :

$$\boxed{\sum_{i=1}^{\infty} f(\mu(1-\mu)^{i-1}b_it) = \frac{1}{\mu} \int_0^{\mu pt} \frac{f(\tau)}{\tau} d\tau + R(t)} \quad (100)$$

where:

$$\boxed{|R(t)| \leq \frac{h_1(\mu pt)}{2-\mu} + \frac{\mathcal{C}\lambda h_2(\mu t)}{1-\lambda(1-\mu)}} \quad (101)$$

*Proof.* First, note that the integral in (100) is well-posed, since the function  $f(t)/t$  is continuous for all  $t \in \mathbb{R}$ . Then, for the case  $t = 0$ , Eq. (100) is trivially verified, since  $f(0) = 0$  by assumption, and  $h_1(0) = h_2(0) = 0$  by definitions (96) and (97). We now prove the lemma for the case  $t > 0$ , and the proof for  $t < 0$  will follow similarly.

Let us introduce the infinite partition of the interval  $[0, \mu pt]$ :

$$\tau_i = \mu(1-\mu)^{i-1}pt, \quad i = 1, 2, \dots \quad (102)$$

By using a first-order Taylor expansion we can write:

$$\begin{aligned} & \sum_{i=1}^n f(\mu(1-\mu)^{i-1}b_it) \\ &= \sum_{i=1}^n f(\tau_i) + \sum_{i=1}^n \mu(1-\mu)^{i-1}f'(t_i)(b_i - p)t. \end{aligned} \quad (103)$$

for a value  $t_i$  that is certainly contained in the interval  $[0, \mu t]$ , because so are the points  $\mu(1 - \mu)^{i-1}b_i t$  and  $\tau_i$ . Let us focus on the first term on the RHS of (103), and let  $g(t) = f(t)/t$ . A second-order Taylor expansion of the function  $G(t) = \int_t^{\tau_i} g(\tau) d\tau$  around the point  $\tau_i$  gives:

$$\begin{aligned} G(\tau_{i+1}) &= G(\tau_i) + g(\tau_i)\delta_i - g'(\bar{t}_i)\frac{\delta_i^2}{2} \\ &= g(\tau_i)\delta_i - g'(\bar{t}_i)\frac{\delta_i^2}{2}, \end{aligned} \quad (104)$$

for a certain  $\bar{t}_i \in (\tau_{i+1}, \tau_i)$  and where

$$\delta_i = \tau_i - \tau_{i+1} = \mu \tau_i. \quad (105)$$

We then obtain from (104) that

$$\begin{aligned} \int_{\mu(1-\mu)^n pt}^{\mu pt} g(\tau) d\tau &= \sum_{i=1}^n \int_{\tau_{i+1}}^{\tau_i} g(\tau) d\tau \\ &= \sum_{i=1}^n g(\tau_i)\delta_i - \sum_{i=1}^n g'(\bar{t}_i)\frac{\delta_i^2}{2} \\ &= \mu \sum_{i=1}^n f(\tau_i) - \sum_{i=1}^n g'(\bar{t}_i)\frac{\delta_i^2}{2}, \end{aligned} \quad (106)$$

Computing now  $\sum_{i=1}^n f(\tau_i)$  from (106), and substituting into (103), we have

$$\sum_{i=1}^n f(\mu(1-\mu)^{i-1}b_i t) = \frac{1}{\mu} \int_{\mu(1-\mu)^n pt}^{\mu pt} \frac{f(\tau)}{\tau} d\tau + r_n(t), \quad (107)$$

where

$$\begin{aligned} r_n(t) &\triangleq \frac{1}{\mu} \sum_{i=1}^n g'(\bar{t}_i)\frac{\delta_i^2}{2} + \\ &\quad \sum_{i=1}^n \mu(1-\mu)^{i-1} f'(\bar{t}_i)(b_i - p)t. \end{aligned} \quad (108)$$

Both series on the RHS of (108) are absolutely convergent as  $n \rightarrow \infty$ . Indeed, using the definition of the auxiliary function  $h_1(t)$  in (96), we see that the first series satisfies:

$$\begin{aligned} \frac{1}{\mu} \sum_{i=1}^n |g'(\bar{t}_i)|\frac{\delta_i^2}{2} &= \frac{\mu}{2} (\mu pt)^2 \sum_{i=1}^n |g'(\bar{t}_i)|(1-\mu)^{2(i-1)} \\ &\leq \mu \sum_{i=1}^{\infty} (1-\mu)^{2(i-1)} h_1(\mu pt) = \frac{h_1(\mu pt)}{2-\mu}. \end{aligned} \quad (109)$$

Similarly, using the definition of  $h_2(t)$  in (97), and the assumption  $|b_i - p| \leq \mathcal{C}\lambda^i$ , we see that the second series obeys:

$$\begin{aligned}
& \sum_{i=1}^n \mu(1-\mu)^{i-1} |f'(t_i)(b_i - p)|t \\
& \leq h_2(\mu t) \sum_{i=1}^n (1-\mu)^{i-1} |b_i - p| \\
& \leq \mathcal{C}\lambda h_2(\mu t) \sum_{i=1}^{\infty} [\lambda(1-\mu)]^{i-1} \\
& = \frac{\mathcal{C}\lambda h_2(\mu t)}{1 - \lambda(1-\mu)}. \tag{110}
\end{aligned}$$

Convergence of the first term on the RHS in (107) follows by the definition of integration. If we now denote by  $R(t)$  the limit of  $r_n(t)$  as  $n \rightarrow \infty$ , we have in fact proved (100) along with the upper bound in (101).

□

The next lemma establishes some useful properties of the function  $\omega(t)$  defined in (30).

**LEMMA 2 (Derivatives of  $\omega(t)$ ).** *Let  $\psi(t)$  be the LMGF of the local statistic  $x_k(n)$ , and assume that  $\psi(t) < \infty$  for all  $t \in \mathbb{R}$ . Then, the function  $\omega(t)$  in (30) is infinitely differentiable in  $\mathbb{R}$ , and, for all  $r \in \mathbb{N}$ :*

$$\boxed{\omega^{(r)}(t) = \frac{1}{t^r} \int_0^t \psi^{(r)}(\tau) \tau^{r-1} d\tau} \tag{111}$$

where, for  $t = 0$ , the above equation should be read as:

$$\omega^{(r)}(0) = \lim_{t \rightarrow 0} \frac{1}{t^r} \int_0^t \psi^{(r)}(\tau) \tau^{r-1} d\tau = \frac{\psi^{(r)}(0)}{r}. \tag{112}$$

*Proof.* We prove the lemma by induction. Let us consider first the case  $t \neq 0$ . Property (111) holds for  $r = 1$ :

$$\omega'(t) = \frac{\psi(t)}{t} = \frac{1}{t} \int_0^t \psi'(\tau) d\tau, \tag{113}$$

having used  $\psi(0) = 0$  since  $\psi(t)$  is a LMGF. Now we show that if (111) holds for some  $r$ , then it must hold for  $r + 1$ . Indeed, note that

$$\omega^{(r+1)}(t) = \frac{d}{dt} \omega^{(r)}(t) = \frac{d}{dt} \left( \frac{1}{t^r} \int_0^t \psi^{(r)}(\tau) \tau^{r-1} d\tau \right). \tag{114}$$

Applying the rule of integration by parts we have:

$$\begin{aligned}
& \frac{1}{t^r} \int_0^t \psi^{(r)}(\tau) \tau^{r-1} d\tau \\
&= \frac{1}{t^r} \left[ \psi^{(r)}(\tau) \frac{\tau^r}{r} \right]_0^t - \frac{1}{t^r} \int_0^t \psi^{(r+1)}(\tau) \frac{\tau^r}{r} d\tau \\
&= \frac{\psi^{(r)}(t)}{r} - \frac{1}{rt^r} \int_0^t \psi^{(r+1)}(\tau) \tau^r d\tau.
\end{aligned} \tag{115}$$

Differentiating the above expression yields:

$$\begin{aligned}
\omega^{(r+1)}(t) &= \frac{\psi^{(r+1)}(t)}{r} - \frac{\psi^{(r+1)}(t)}{r} \\
&+ \frac{1}{t^{r+1}} \int_0^t \psi^{(r+1)}(\tau) \tau^r d\tau.
\end{aligned} \tag{116}$$

This proves the claim for  $t \neq 0$ . Now, since  $\omega^{(r)}(t)$  exists for all  $t \neq 0$ , we have [43]:

$$\omega^{(r)}(0) = \lim_{t \rightarrow 0} \omega^{(r)}(t), \tag{117}$$

provided that the above limit exists. By applying L'Hospital's rule [43] we have in fact:

$$\omega^{(r)}(0) = \lim_{t \rightarrow 0} \frac{1}{t^r} \int_0^t \psi^{(r)}(\tau) \tau^{r-1} d\tau = \frac{\psi^{(r)}(0)}{r} \tag{118}$$

and the proof is complete. □

We are now ready to prove Theorem 1.

*Proof of Theorem 1.* Recalling the definition of  $\xi_{i,\ell}$  in (87), the finite-horizon random variable  $\mathbf{y}_k^*(n)$  in (21), and its LMGF  $\phi_{k,\mu}(t; n)$  in (58) can be written as, respectively:

$$\mathbf{y}_k^*(n) \triangleq \sum_{i=1}^n \sum_{\ell=1}^S \xi_{i,\ell} \mathbf{x}_\ell(i), \tag{119}$$

and

$$\phi_{k,\mu}(t; n) = \sum_{i=1}^n \sum_{\ell=1}^S \psi(\xi_{i,\ell} t) = \sum_{\ell=1}^S \sum_{i=1}^n \psi(\xi_{i,\ell} t). \tag{120}$$

We now apply Lemma 1 to each of the  $S$  inner summations, with the choices  $f(t) = \psi(t)$ ,  $b_i = b_{k,\ell}(i)$ , and  $p = p_\ell$ . Note that the hypotheses of the lemma are satisfied. Indeed  $\psi(t)$  is a LMGF, implying that  $\psi(0) = 0$  and that  $\psi(t)$  is infinitely differentiable in  $\mathbb{R}$ . By assumption, we

have  $\psi(t) < \infty$  for all  $t \in \mathbb{R}$ . As to the conditions in (99), by Perron's theorem [36, Theorem 8.5.1], there exist  $\mathcal{C} > 0$  and  $0 < \lambda < 1$  such that:

$$|b_{k,\ell}(i) - p_\ell| \leq \mathcal{C}\lambda^i. \quad (121)$$

By Lemma 1 we conclude that

$$\begin{aligned} \lim_{n \rightarrow \infty} \phi_{k,\mu}(t; n) &= \sum_{i=1}^{\infty} \sum_{\ell=1}^S \psi(\xi_{i,\ell} t) \\ &= \frac{1}{\mu} \sum_{\ell=1}^S \int_0^{p_\ell \mu t} \frac{\psi(\tau)}{\tau} d\tau + \sum_{\ell=1}^S R_\ell(t), \end{aligned} \quad (122)$$

where we made explicit the dependence of the remainder terms upon  $\ell = 1, 2, \dots, S$ .

In view of the continuity theorem for the moment generating functions [45], and since by definition  $\mathbf{y}_k^*(n)$  converges in distribution to  $\mathbf{y}_{k,\mu}^*$ , the limit of  $\phi_{k,\mu}(t; n)$  (i.e., the limit of the LMGF of  $\mathbf{y}_k^*(n)$ ) represents  $\phi_{k,\mu}(t)$  (i.e., the LMGF of  $\mathbf{y}_{k,\mu}^*$ ). Formally, we have:

$$\phi_{k,\mu}(t) = \lim_{n \rightarrow \infty} \phi_{k,\mu}(t; n) = \sum_{i=1}^{\infty} \sum_{\ell=1}^S \psi(\xi_{i,\ell} t), \quad (123)$$

and claim (88) is proved. For the same reason we can rewrite (122) as:

$$\phi_{k,\mu}(t) = \frac{1}{\mu} \sum_{\ell=1}^S \int_0^{p_\ell \mu t} \frac{\psi(\tau)}{\tau} d\tau + \sum_{\ell=1}^S R_\ell(t), \quad (124)$$

and

$$\mu \phi_{k,\mu}(t/\mu) = \sum_{\ell=1}^S \int_0^{p_\ell t} \frac{\psi(\tau)}{\tau} d\tau + \mu \sum_{\ell=1}^S R_\ell(t/\mu), \quad (125)$$

where, from (101), the remainder term is bounded by:

$$\mu \sum_{\ell=1}^S |R_\ell(t/\mu)| \leq \frac{\mu}{2-\mu} \sum_{\ell=1}^S h_1(p_\ell t) + \frac{\mu S \mathcal{C} \lambda h_2(t)}{1-\lambda(1-\mu)}. \quad (126)$$

We conclude from (124) that claim (89) and claim (93) for the case  $r = 0$  hold.

To establish the remaining claims, we start by noting from (120) that, for all  $r = 1, 2, \dots$ :

$$\phi_{k,\mu}^{(r)}(t; n) = \sum_{i=1}^n \sum_{\ell=1}^S \xi_{i,\ell}^r \psi^{(r)}(\xi_{i,\ell} t). \quad (127)$$

Now, recalling the definition of  $\xi_{i,\ell}$  from (87), we have:

$$0 \leq \xi_{i,\ell} \leq \mu(1-\mu)^{i-1}, \quad t \in [-a, a] \Rightarrow \xi_{i,\ell} t \in [-a, a]. \quad (128)$$

Thus, for all  $t \in [-a, a]$  we can write:

$$|\xi_{i,\ell}^r \psi^{(r)}(\xi_{i,\ell} t)| \leq \mu^r (1 - \mu)^{r(i-1)} \max_{\tau \in [-a, a]} |\psi^{(r)}(\tau)|. \quad (129)$$

But since

$$\mu^r \sum_{i=1}^{\infty} (1 - \mu)^{r(i-1)} = \frac{\mu^r}{1 - (1 - \mu)^r}, \quad (130)$$

we conclude that  $\phi_{k,\mu}^{(r)}(t; n)$  is majorized by an absolutely convergent series. In view of Weierstrass theorem about uniform convergence [43, Theorem 7.10, p. 148], this result implies that  $\phi_{k,\mu}^{(r)}(t; n)$  converges uniformly as  $n \rightarrow \infty$  in any compact interval  $[-a, a]$ . Since this holds for all  $r = 1, 2, \dots$ , and since we already showed that  $\phi_{k,\mu}(t; n) \rightarrow \phi_{k,\mu}(t)$  in (88), the results about uniform convergence and differentiability [43, Theorem 7.17, p. 152] allow us to conclude that the derivative of the limit function equals the limit of the series of derivatives, namely, that (90) holds true. Moreover, using (127) for  $n \rightarrow \infty$ , we can write for any  $t \neq 0$ :

$$\phi_{k,\mu}^{(r)}(t) = \frac{1}{t^r} \sum_{i=1}^{\infty} \sum_{\ell=1}^S (\xi_{i,\ell} t)^r \psi^{(r)}(\xi_{i,\ell} t). \quad (131)$$

Applying Lemma 1 to the function  $f(t) = t^r \psi^{(r)}(t)$ , with the same choices done before for  $b_i$  and  $p$ , Eq. (100) gives:

$$\phi_{k,\mu}^{(r)}(t) = \frac{1}{\mu t^r} \sum_{\ell=1}^S \int_0^{p_\ell \mu t} \psi^{(r)}(\tau) \tau^{r-1} d\tau + \frac{1}{t^r} \sum_{\ell=1}^S R_\ell(t), \quad (132)$$

which implies:

$$\frac{\phi_{k,\mu}^{(r)}(t/\mu)}{\mu^{r-1}} = \frac{1}{t^r} \sum_{\ell=1}^S \int_0^{p_\ell t} \psi^{(r)}(\tau) \tau^{r-1} d\tau + \frac{\mu}{t^r} \sum_{\ell=1}^S R_\ell(t/\mu), \quad (133)$$

and, using (101), the remainder term vanishes as  $\mu \rightarrow 0$ , implying that

$$\frac{\phi_{k,\mu}^{(r)}(t/\mu)}{\mu^{r-1}} \xrightarrow{\mu \rightarrow 0} \frac{1}{t^r} \sum_{\ell=1}^S \int_0^{p_\ell t} \psi^{(r)}(\tau) \tau^{r-1} d\tau. \quad (134)$$

Now, in view of Lemma 2, and by the definition of  $\phi(t)$  in (89), we have:

$$\phi^{(r)}(t) = \sum_{\ell=1}^S p_\ell^r \omega^{(r)}(p_\ell t) = \frac{1}{t^r} \sum_{\ell=1}^S \int_0^{p_\ell t} \psi^{(r)}(\tau) \tau^{r-1} d\tau. \quad (135)$$

Comparing (134) and (135) we arrive at (91) for all  $t \neq 0$ . It remains to consider the case  $t = 0$ , namely, to show that:

$$\frac{\phi_{k,\mu}^{(r)}(0)}{\mu^{r-1}} \xrightarrow{\mu \rightarrow 0} \phi^{(r)}(0) = \frac{\psi^{(r)}(0)}{r} \sum_{\ell=1}^S p_\ell^r, \quad (136)$$

where the last equality follows from (112). This equality is established as follows:

$$\begin{aligned}
\phi_{k,\mu}^{(r)}(0) &\stackrel{(90)}{=} \psi^{(r)}(0) \sum_{i=1}^{\infty} \sum_{\ell=1}^S \xi_{i,\ell}^r \\
&\stackrel{(87)}{=} \psi^{(r)}(0) \mu^r \sum_{i=1}^{\infty} \sum_{\ell=1}^S (1-\mu)^{r(i-1)} b_{k,\ell}^r(i) \\
&= \psi^{(r)}(0) \mu^r \sum_{i=1}^{\infty} \sum_{\ell=1}^S p_{\ell}^r (1-\mu)^{r(i-1)} + \\
&\quad \underbrace{\psi^{(r)}(0) \mu^r \sum_{i=1}^{\infty} \sum_{\ell=1}^S (1-\mu)^{r(i-1)} (b_{k,\ell}^r(i) - p_{\ell}^r)}_{\triangleq R} \\
&= \psi^{(r)}(0) \frac{\mu^r}{1 - (1-\mu)^r} \sum_{\ell=1}^S p_{\ell}^r + R,
\end{aligned} \tag{137}$$

or

$$\frac{\phi_{k,\mu}^{(r)}(0)}{\mu^{r-1}} = \psi^{(r)}(0) \frac{\mu}{1 - (1-\mu)^r} \sum_{\ell=1}^S p_{\ell}^r + \frac{R}{\mu^{r-1}}. \tag{138}$$

Using the inequality<sup>1</sup>

$$|b_{k,\ell}^r(i) - p_{\ell}^r| \leq r |b_{k,\ell}(i) - p_{\ell}|, \tag{139}$$

and (121) we can write

$$\frac{|R|}{\mu^{r-1}} \leq r SC \lambda \psi^{(r)}(0) \frac{\mu}{1 - [\lambda(1-\mu)^r]}. \tag{140}$$

Therefore, the second term on the RHS of (138) vanishes as  $\mu \rightarrow 0$ , and by application of L'Hospital's rule [43] to the first term we get (136). Finally, examining (133) and (140) shows that the  $\mathcal{O}(\mu)$  error estimate in (93) holds true.

<sup>1</sup>The inequality follows from the factorization, holding for  $r \in \mathbb{N}$ :

$$a^r - b^r = (a - b) \sum_{k=0}^{r-1} a^k b^{r-1-k},$$

and from the fact that  $b_{k,\ell}(i)$  and  $p_{\ell}$  are not greater than one.



## APPENDIX B

The proof relies on some properties of moment generating functions, which are collected in Sec. B-A, and on a technical lemma, whose proof is deferred to Sec. B-B.

*Proof of Theorem 3.* By definition,  $\theta_\gamma$  solves the stationary equation

$$\phi'(\theta_\gamma) = \gamma. \quad (141)$$

Since  $\phi''(t) > 0$ , it follows that  $\phi'(t)$  is strictly increasing, and, hence,  $\theta_\gamma$  is unique. Moreover, from (92) used with  $r = 1$  we obtain:

$$\phi'(0) = \psi'(0) = \mathbb{E}[\mathbf{x}], \quad (142)$$

recalling that the first derivative of the LMGF, evaluated at the origin, gives the expectation of the random variable [34], [35]. The equality  $\phi'(0) = \mathbb{E}[\mathbf{x}]$  implies that  $\theta_\gamma > 0$ , because by assumption  $\gamma > \mathbb{E}[\mathbf{x}]$ , and  $\phi'(t)$  is strictly increasing.

For ease of notation, in the following we shall suppress the subscript  $\gamma$ , and we shall use the symbol  $\theta$  to denote the unique solution of (141). Let us denote by  $m(dy)$  the probability measure associated with  $\mathbf{y}_{k,\mu}^*$ , and introduce the following measure transformation, namely, an exponential translation in the form:

$$\tilde{m}(dy) = e^{\frac{\theta}{\mu} y - \phi_{k,\mu}\left(\frac{\theta}{\mu}\right)} m(dy). \quad (143)$$

The use of a similar transformation for large deviations analysis was originally proposed by Cramér [41], and then used, e.g., in [33], to find the exact asymptotics in the simplest case of normalized sums of i.i.d. random variables — see also [34], [35]. We will explain how this transformation is also helpful in our more general case, where the asymptotic parameter  $\mu$  is not simply the index of a summation of i.i.d. random variables.

Denoting by  $\mathcal{I}_{\mathcal{E}}$  the indicator of an event  $\mathcal{E}$ , and by  $\mathbb{E}_{\tilde{m}}[\cdot]$  the expectation operator under the

measure  $\tilde{m}(dy)$ , some straightforward algebra allows us to write:

$$\begin{aligned}
\mathbb{P}[\mathbf{y}_{k,\mu}^* > \gamma] &= \mathbb{E}[\mathcal{I}_{\{\mathbf{y}_{k,\mu}^* > \gamma\}}] = \int_{\gamma}^{\infty} m(dy) \\
&= \int_{\gamma}^{\infty} e^{-\frac{\theta}{\mu} y + \phi_{k,\mu}(\frac{\theta}{\mu})} \tilde{m}(dy) \\
&= \mathbb{E}_{\tilde{m}}[e^{-\frac{\theta}{\mu} \mathbf{y}_{k,\mu}^* + \phi_{k,\mu}(\frac{\theta}{\mu})} \mathcal{I}_{\{\mathbf{y}_{k,\mu}^* > \gamma\}}] \\
&= e^{-\frac{1}{\mu}[\gamma\theta - \phi(\theta)]} e^{-\frac{1}{\mu}[\phi(\theta) - \mu\phi_{k,\mu}(\frac{\theta}{\mu})]} e^{\frac{\theta}{\mu}[\gamma - \phi'_{k,\mu}(\frac{\theta}{\mu})]} \times \\
&\quad \mathbb{E}_{\tilde{m}}[e^{-\frac{\theta}{\mu}[\mathbf{y}_{k,\mu}^* - \phi'_{k,\mu}(\frac{\theta}{\mu})]} \mathcal{I}_{\{\mathbf{y}_{k,\mu}^* > \gamma\}}] \\
&= e^{-\frac{1}{\mu}[\Phi(\gamma) + \epsilon_{k,\mu}(\theta)]} e^{\frac{\theta}{\mu}[\gamma - \phi'_{k,\mu}(\frac{\theta}{\mu})]} \times \\
&\quad \mathbb{E}_{\tilde{m}}[e^{-\frac{\theta}{\mu}[\mathbf{y}_{k,\mu}^* - \phi'_{k,\mu}(\frac{\theta}{\mu})]} \mathcal{I}_{\{\mathbf{y}_{k,\mu}^* > \gamma\}}], \tag{144}
\end{aligned}$$

where we used definition (51a) for  $\epsilon_{k,\mu}(\theta)$ , and the fact that

$$\Phi(\gamma) = \sup_{t \in \mathbb{R}} [\gamma t - \phi(t)] \Leftrightarrow \Phi(\gamma) = \gamma\theta - \phi(\theta), \tag{145}$$

which holds true because  $\phi(t)$  is strictly convex and  $\theta$  solves the stationary equation (141). It is now useful to introduce the normalized random variable

$$\mathbf{w}_{k,\mu}^* \triangleq \frac{\mathbf{y}_{k,\mu}^* - \phi'_{k,\mu}(\theta/\mu)}{\sqrt{\phi''_{k,\mu}(\theta/\mu)}}. \tag{146}$$

For ease of notation, we introduce the quantity

$$s_{k,\mu} \triangleq \sqrt{\frac{\phi''_{k,\mu}(\theta/\mu)}{\mu}} \xrightarrow{\mu \rightarrow 0} \sqrt{\phi''(\theta)}, \tag{147}$$

where the convergence follows from (91). With this definition, the expression for  $\mathbf{w}_{k,\mu}^*$  can be rewritten as:

$$\mathbf{w}_{k,\mu}^* = \frac{\mathbf{y}_{k,\mu}^* - \phi'_{k,\mu}(\theta/\mu)}{\sqrt{\mu} s_{k,\mu}}. \tag{148}$$

It is also convenient to introduce the normalized threshold:

$$\gamma_{k,\mu} \triangleq \frac{\gamma - \phi'_{k,\mu}(\theta/\mu)}{\sqrt{\mu} s_{k,\mu}}, \tag{149}$$

and the normalized expectation:

$$C_{k,\mu} \triangleq \frac{1}{\sqrt{\mu}} \mathbb{E}_{\tilde{m}} \left[ e^{-\frac{\theta s_{k,\mu}}{\sqrt{\mu}} (\mathbf{w}_{k,\mu}^* - \gamma_{k,\mu})} \mathcal{I}_{\{\mathbf{w}_{k,\mu}^* > \gamma_{k,\mu}\}} \right]. \tag{150}$$

Using the above definitions, Eq. (144) can be rewritten as:

$$\mathbb{P}[\mathbf{y}_{k,\mu}^* > \gamma] = \sqrt{\mu} C_{k,\mu} e^{-\frac{1}{\mu}[\Phi(\gamma) + \epsilon_{k,\mu}(\theta)]}. \quad (151)$$

Examining expression (50) for  $\mathcal{P}_{k,\mu}(\gamma)$ , it is immediate to recognize that the theorem will be established if we are able to show that:

$$\boxed{C_{k,\mu} \xrightarrow{\mu \rightarrow 0} \frac{1}{\sqrt{2\pi\theta^2\phi''(\theta)}}} \quad (152)$$

The key for proving the above claim is the asymptotic normality of the random variable  $\mathbf{w}_{k,\mu}^*$  under the transformed measure  $\tilde{m}(dy)$ . To see this, let us exploit the fundamental properties of  $\mathbf{w}_{k,\mu}^*$ . First, according to property P2 in Sec. B-A, and since the first three central moments of a random variable coincide with its first three cumulants [48], under the considered transformed measure we have:

$$\mathbb{E}_{\tilde{m}}[\mathbf{y}_{k,\mu}^*] = \phi'_{k,\mu}(\theta/\mu), \quad (153)$$

$$\mathbb{E}_{\tilde{m}}[(\mathbf{y}_{k,\mu}^* - \phi'_{k,\mu}(\theta/\mu))^2] = \phi''_{k,\mu}(\theta/\mu), \quad (154)$$

$$\mathbb{E}_{\tilde{m}}[(\mathbf{y}_{k,\mu}^* - \phi'_{k,\mu}(\theta/\mu))^3] = \phi'''_{k,\mu}(\theta/\mu). \quad (155)$$

Accordingly, from (146) and (147), we conclude that  $\mathbf{w}_{k,\mu}^*$  is zero-mean, unit-variance, and its third moment is:

$$\mathbb{E}_{\tilde{m}}[(\mathbf{w}_{k,\mu}^*)^3] = \frac{\phi'''_{k,\mu}(\theta/\mu)}{\mu^{3/2}s_{k,\mu}^3}. \quad (156)$$

A crucial step consists now in finding an asymptotic expansion for the cumulative distribution function of the random variable  $\mathbf{w}_{k,\mu}^*$  under the transformed measure. To this aim, we follow the procedure employed for deriving the Berry-Esseen and Edgeworth expansions — see, e.g., [40], and we introduce the function:

$$\begin{aligned} G_{k,\mu}(w) &\triangleq N(w) - \frac{\mathbb{E}_{\tilde{m}}[(\mathbf{w}_{k,\mu}^*)^3]}{6}(w^2 - 1)n(w) \\ &= N(w) - \frac{\phi'''_{k,\mu}(\theta/\mu)}{6\mu^{3/2}s_{k,\mu}^3}(w^2 - 1)n(w), \end{aligned} \quad (157)$$

where  $N(w)$  and  $n(w) = N'(w)$  are, respectively, the cumulative distribution function and the probability density function of a standard normal. Denoting by  $F_{k,\mu}(w)$  the cumulative

distribution function of the random variable  $w_{k,\mu}^*$  under the transformed measure, in B-B we establish that (Lemma 3):

$$\frac{1}{\sqrt{\mu}} \sup_{w \in \mathbb{R}} |F_{k,\mu}(w) - G_{k,\mu}(w)| \xrightarrow{\mu \rightarrow 0} 0. \quad (158)$$

This implies that  $w_{k,\mu}^*$  is asymptotically normal under the transformed measure  $\tilde{m}(dy)$ . More importantly, Eq. (158) provides a refined, uniform estimate of the convergence error, a property that we shall exploit soon to prove (152). To this aim, we evaluate explicitly the expectation in (150). Using integration by parts, we obtain:

$$\begin{aligned} C_{k,\mu} &= \frac{1}{\sqrt{\mu}} \int_{\gamma_{k,\mu}}^{\infty} e^{-\frac{\theta s_{k,\mu}}{\sqrt{\mu}}(w-\gamma_{k,\mu})} dF_{k,\mu}(w) \\ &= \frac{1}{\sqrt{\mu}} \left[ e^{-\frac{\theta s_{k,\mu}}{\sqrt{\mu}}(w-\gamma_{k,\mu})} [F_{k,\mu}(w) - F_{k,\mu}(\gamma_{k,\mu})] \right]_{\gamma_{k,\mu}}^{\infty} + \\ &\quad \frac{\theta s_{k,\mu}}{\mu} \int_{\gamma_{k,\mu}}^{\infty} e^{-\frac{\theta s_{k,\mu}}{\sqrt{\mu}}(w-\gamma_{k,\mu})} [F_{k,\mu}(w) - F_{k,\mu}(\gamma_{k,\mu})] dw \\ &= \frac{1}{\sqrt{\mu}} \int_0^{\infty} e^{-x} [F_{k,\mu}(p_{k,\mu}(x)) - F_{k,\mu}(\gamma_{k,\mu})] dx, \end{aligned} \quad (159)$$

where the last equality follows from the change of variable

$$\frac{\theta s_{k,\mu}}{\sqrt{\mu}}(w - \gamma_{k,\mu}) \rightarrow x, \quad (160)$$

and from the definition:

$$p_{k,\mu}(x) \triangleq \frac{\sqrt{\mu}}{\theta s_{k,\mu}} x + \gamma_{k,\mu}. \quad (161)$$

Let us also introduce the analogue of (159) for the normal approximation  $G_{k,\mu}(w)$  in (157), namely,

$$D_{k,\mu} \triangleq \frac{1}{\sqrt{\mu}} \int_0^{\infty} e^{-x} [G_{k,\mu}(p_{k,\mu}(x)) - G_{k,\mu}(\gamma_{k,\mu})] dx. \quad (162)$$

By using Lemma 3, we can write:

$$\begin{aligned} |C_{k,\mu} - D_{k,\mu}| &\leq \frac{2}{\sqrt{\mu}} \sup_{w \in \mathbb{R}} |F_{k,\mu}(w) - G_{k,\mu}(w)| \int_0^{\infty} e^{-x} dx \\ &= \frac{2}{\sqrt{\mu}} \sup_{w \in \mathbb{R}} |F_{k,\mu}(w) - G_{k,\mu}(w)| \xrightarrow{\mu \rightarrow 0} 0. \end{aligned} \quad (163)$$

By virtue of this result, and in order to prove (152), it is enough to show that

$$D_{k,\mu} \xrightarrow{\mu \rightarrow 0} \frac{1}{\sqrt{2\pi\theta^2\phi''(\theta)}}. \quad (164)$$

Substituting definition (157) for  $G_{k,\mu}(w)$  into (162), we have:

$$\begin{aligned}
 D_{k,\mu} = & \frac{1}{\sqrt{\mu}} \int_0^\infty e^{-x} [N(p_{k,\mu}(x)) - N(\gamma_{k,\mu})] dx + \\
 & \frac{\phi_{k,\mu}'''(\theta/\mu)}{6\mu^2 s_{k,\mu}^3} \int_0^\infty e^{-x} \times \\
 & [n(\gamma_{k,\mu})(\gamma_{k,\mu}^2 - 1) - n(p_{k,\mu}(x))(p_{k,\mu}^2(x) - 1)] dx.
 \end{aligned} \tag{165}$$

Consider first the second term on the RHS. By (93) we can write (recall from (141) that  $\phi'(\theta) = \gamma$ ):

$$\phi'_{k,\mu}(\theta/\mu) = \phi'(\theta) + \mathcal{O}(\mu) = \gamma + \mathcal{O}(\mu), \tag{166}$$

which, in view of (149), implies that

$$\lim_{\mu \rightarrow 0} \gamma_{k,\mu} = 0. \tag{167}$$

We therefore conclude that  $\lim_{\mu \rightarrow 0} p_{k,\mu}(x) = 0$  from (161) and (147). Accordingly, we have:

$$\lim_{\mu \rightarrow 0} [n(\gamma_{k,\mu})(\gamma_{k,\mu}^2 - 1) - n(p_{k,\mu}(x))(p_{k,\mu}^2(x) - 1)] = 0, \tag{168}$$

and, hence, by dominated convergence [46] (note that the function  $n(w)(w^2 - 1)$  is bounded), the second term on the RHS in (165) vanishes as  $\mu \rightarrow 0$ , having further used the fact that the ratio  $\frac{\phi_{k,\mu}'''(\theta/\mu)}{6\mu^2 s_{k,\mu}^3}$  converges in view of (91) and (147).

With regards to the first term on the RHS of (165), by using a second-order Taylor expansion of  $N(w)$  around the point  $\gamma_{k,\mu}$  we obtain, for an intermediate point  $q_{k,\mu}(x)$ :

$$\begin{aligned}
 N(p_{k,\mu}(x)) &= N(\gamma_{k,\mu}) + n(\gamma_{k,\mu})(p_{k,\mu}(x) - \gamma_{k,\mu}) + \\
 &\quad n'(q_{k,\mu}(x)) \frac{(p_{k,\mu}(x) - \gamma_{k,\mu})^2}{2} \\
 &= N(\gamma_{k,\mu}) + n(\gamma_{k,\mu}) \frac{\sqrt{\mu}}{\theta s_{k,\mu}} x + \\
 &\quad n'(q_{k,\mu}(x)) \frac{\mu x^2}{2(\theta s_{k,\mu})^2},
 \end{aligned} \tag{169}$$

where we used (161). The above expansion then gives:

$$\begin{aligned}
& \frac{1}{\sqrt{\mu}} \int_0^\infty e^{-x} [N(p_{k,\mu}(x)) - N(\gamma_{k,\mu})] dx \\
&= \frac{n(\gamma_{k,\mu})}{\theta s_{k,\mu}} \int_0^\infty x e^{-x} dx + \\
& \quad \frac{\sqrt{\mu}}{2(\theta s_{k,\mu})^2} \int_0^\infty x^2 e^{-x} n'(q_{k,\mu}(x)) dx.
\end{aligned} \tag{170}$$

The second term on the RHS of (170) vanishes because

$$\begin{aligned}
\left| \int_0^\infty x^2 e^{-x} n'(q_{k,\mu}(x)) dx \right| &\leq \sup_{w \in \mathbb{R}} |n'(w)| \int_0^\infty x^2 e^{-x} dx \\
&= 2 \sup_{w \in \mathbb{R}} |n'(w)| < \infty.
\end{aligned} \tag{171}$$

On the other hand, the first term on the RHS of (170) satisfies:

$$\frac{n(\gamma_{k,\mu})}{\theta s_{k,\mu}} \int_0^\infty x e^{-x} dx = \frac{n(\gamma_{k,\mu})}{\theta s_{k,\mu}} \xrightarrow{\mu \rightarrow 0} \frac{1}{\sqrt{2\pi\theta^2\phi''(\theta)}}, \tag{172}$$

since  $\gamma_{k,\mu}$  vanishes by (167),  $n(0) = 1/\sqrt{2\pi}$ ,  $\int_0^\infty x e^{-x} dx = 1$ , and  $s_{k,\mu}$  converges to  $\sqrt{\phi''(\theta)}$  by (147). The proof of (51a) is now complete.

It remains to justify the alternative expression (51b). The rationale behind our choice is as follows. Examining (172), it is seen that the term  $n(\gamma_{k,\mu})$  is replaced by its limiting constant value of  $1/\sqrt{2\pi}$ . In order to get a better approximation for finite values of  $1/\mu$ , we can decide to retain the dependence on  $\mu$  by including  $n(\gamma_{k,\mu})$  in the main formula. This amounts to replacing  $C_{k,\mu}$  in (151), not with the limiting value in (152), but rather with

$$\sqrt{2\pi} n(\gamma_{k,\mu}) \frac{1}{\sqrt{2\pi\theta^2\phi''(\theta)}}. \tag{173}$$

By making explicit the definition of  $\gamma_{k,\mu}$  in (149), and of the Gaussian pdf, the additional term is:

$$\sqrt{2\pi} n(\gamma_{k,\mu}) = e^{-\frac{[\gamma - \phi'_{k,\mu}(\theta/\mu)]^2}{2\mu s_{k,\mu}^2}} \sim e^{-\frac{1}{\mu} \frac{[\phi'(\theta) - \phi'_{k,\mu}(\theta/\mu)]^2}{2\phi''(\theta)}}, \tag{174}$$

where, in the last step, we have used  $\gamma = \phi'(\theta)$ , and, to be consistent with the approximations adopted so far, we have replaced  $s_{k,\mu}^2$  with its limiting value  $\phi''(\theta)$ . Examining the correction at the exponent of (174) we recognize exactly the second term in (51b).

□

### A. Some general properties of moment generating functions

Consider a random variable  $\mathbf{x}$  with Moment Generating Function (MGF), Logarithmic Moment Generating Function (LMGF), and Characteristic Function (CF) defined respectively, as:

$$\nu(t) \triangleq \mathbb{E}[e^{t\mathbf{x}}], \quad \psi(t) \triangleq \ln \nu(t), \quad \varphi(u) \triangleq \mathbb{E}[e^{ju\mathbf{x}}]. \quad (175)$$

We stress that, in this specific subsection,  $\mathbf{x}$  is intended to be a *generic* random variable, and there is no need here to give it any particular meaning in terms of a local or a global network statistic. We assume that  $\nu(t)$  exists for all  $t \in \mathbb{R}$ . Under this assumption, we can also define the MGF relative to a complex argument  $z$ , namely:

$$\nu(z) \triangleq \mathbb{E}[e^{z\mathbf{x}}], \quad z \in \mathbb{C}, \quad (176)$$

which is the analytic continuation of  $\nu(t)$  over the complex plane [46]. Note that, for  $z = t + ju$ :

$$|\nu(z)| = |\mathbb{E}[e^{z\mathbf{x}}]| \leq \mathbb{E}[|e^{z\mathbf{x}}|] = \nu(t). \quad (177)$$

In this section, we focus on the following measure transformation (exponential measure translation):

$$\tilde{m}(dx) = e^{\eta x - \psi(\eta)} m(dx) = \frac{e^{\eta x}}{\nu(\eta)} m(dx), \quad (178)$$

and establish several important properties characterizing the statistical behavior of  $\mathbf{x}$  under the transformed measure  $\tilde{m}(dx)$ .

- *Property P1.* The MGF under the transformation is:

$$\boxed{\tilde{\nu}(t) = \frac{\nu(\eta + t)}{\nu(\eta)}} \quad (179)$$

Indeed, by definition:

$$\tilde{\nu}(t) \triangleq \mathbb{E}_{\tilde{m}}[e^{t\mathbf{x}}] = \frac{\mathbb{E}[e^{(\eta+t)\mathbf{x}}]}{\nu(\eta)} = \frac{\nu(\eta + t)}{\nu(\eta)}. \quad (180)$$

- *Property P2.* The  $r$ -th cumulant  $\tilde{\chi}^{(r)}$  of the variable  $\mathbf{x}$  under the transformation is

$$\boxed{\tilde{\chi}^{(r)} = \psi^{(r)}(\eta)} \quad (181)$$

Again, from property P1, the LMGF of  $\mathbf{x}$  under the transformed measure is:

$$\tilde{\psi}(t) \triangleq \ln \tilde{\nu}(t) = \ln \nu(\eta + t) - \ln \nu(\eta) = \psi(\eta + t) - \psi(\eta). \quad (182)$$

Accordingly, the  $r$ -th cumulant of the variable  $\mathbf{x}$  under the transformation is

$$\tilde{\chi}^{(r)} = \left. \frac{d^r}{dt^r} \tilde{\psi}(t) \right|_{t=0} = \psi^{(r)}(\eta). \quad (183)$$

- *Property P3.* The CF of the random variable  $\mathbf{x}$  under the transformation is

$$\tilde{\varphi}(u) = \frac{\nu(\eta + ju)}{\nu(\eta)} \quad (184)$$

where  $\nu(z)$ , for a complex  $z$ , was defined by (180). This result follows from an argument similar to the proof of Property P1.

- *Property P4.* Consider the shifted random variable

$$\hat{\mathbf{x}} = \mathbf{x} - \psi'(\eta). \quad (185)$$

Since the CF of the shifted-and-scaled random variable  $a\mathbf{x} + b$  is

$$\mathbb{E}[e^{jt(ax+b)}] = \varphi(at)e^{jbt}, \quad (186)$$

the CF of the shifted variable  $\hat{\mathbf{x}}$  under the transformed measure is readily computed as:

$$\hat{\varphi}(u) = \tilde{\varphi}(u)e^{-j\psi'(\eta)u} = \frac{\nu(\eta + ju)}{\nu(\eta)} e^{-j\psi'(\eta)u}. \quad (187)$$

Assume that the parameter  $\eta$  ruling the measure transformation lies in a bounded interval, say,  $|\eta| \leq \eta_{\max}$ . Then, for sufficiently small  $\delta > 0$ , a positive constant  $M = M(\delta, \eta_{\max})$  exists such that for  $|u| \leq \delta$ , and for any choice of  $|\eta| \leq \eta_{\max}$ , the following expansion holds for the CF of the shifted random variable  $\hat{\mathbf{x}}$ :

$$\log \hat{\varphi}(u) = -\psi''(\eta) \frac{u^2}{2} + \psi'''(\eta) \frac{(ju)^3}{6} + R(u) \quad (188)$$

with the remainder bounded as:

$$|R(u)| \leq M \frac{u^4}{24}. \quad (189)$$

In order to prove the above property, let us introduce the logarithm of a complex number  $z$ :

$$\log z = \ln |z| + j \arg(z). \quad (190)$$

For  $|z| < 1$ , it can be alternatively defined by the Taylor series [46]:

$$\log(1 + z) = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} z^n. \quad (191)$$



We shall now consider conditions under which the function  $\log \hat{\varphi}(u)$  can be represented by a Taylor series.

Since  $\hat{\varphi}(u)$  is the CF of a zero-mean random variable, the following known bound is obtained from the Taylor expansion of the complex exponential [40, Eq. (4.14), p. 514]:

$$|\hat{\varphi}(u) - 1| \leq \mathbb{E}_{\tilde{m}}[\hat{\mathbf{x}}^2] \frac{u^2}{2} = \psi''(\eta) \frac{u^2}{2}, \quad (192)$$

where in the last equality we used property P2. We now select a parameter  $\delta > 0$  such that

$$\delta^2 < \frac{2}{\max_{|\eta| \leq \eta_{\max}} \psi''(\eta)}, \quad (193)$$

where the above choice is meaningful because  $\psi(\eta)$  is a LMGF, which implies that  $\psi''(\eta) > 0$ , and that  $\psi''(\eta)$  is a continuous function, and thus admits a maximum within the interval  $[-\eta_{\max}, \eta_{\max}]$ . Therefore, over the range  $|u| \leq \delta$ , Eqs. (192) and (193) give:

$$|\hat{\varphi}(u) - 1| < 1. \quad (194)$$

This allows representing, in the considered range for  $u$ , the function  $\log \hat{\varphi}(u)$  via its Taylor series. By the definition of cumulants, this series can be written as [48, Eq. (6.8), p. 45]:

$$\log \hat{\varphi}(u) = \sum_{r=2}^n \tilde{\chi}^{(r)} \frac{(ju)^r}{r!} + o(|u|^n), \quad (u \rightarrow 0). \quad (195)$$

Note that, in the above formula, the series coefficients should be given by the cumulants of  $\hat{\mathbf{x}}$  (computed under the transformed measure). As it can be seen, we used instead the cumulants  $\tilde{\chi}^{(r)}$  of  $\mathbf{x}$ . This is because  $\hat{\mathbf{x}}$  is a shifted version of  $\mathbf{x}$ , and so its cumulants are (but for the first one, which is zero) the same as those of  $\mathbf{x}$ .

Consider now the three-term series. Using the explicit formulas for the cumulants under the transformed measure provided by (181), we get:

$$\log \hat{\varphi}(u) = -\psi''(\eta) \frac{u^2}{2} + \psi'''(\eta) \frac{(ju)^3}{6} + r(\bar{u}) \frac{u^4}{24}, \quad (196)$$

where  $\bar{u} \in (-\delta, \delta)$ , and the function  $r(u)$  can be expressed as:

$$\begin{aligned} r(u) &= \frac{d^4}{du^4} \log \hat{\varphi}(u) \\ &= \frac{d^4}{du^4} \log \left[ \frac{\nu(\eta + ju)}{\nu(\eta)} e^{-j\psi'(\eta)u} \right], \end{aligned} \quad (197)$$

which leads to (the argument  $\eta + ju$  is omitted for ease of notation):

$$r(u) = \frac{1}{\nu^4} \left[ -6(\nu')^4 + 12\nu(\nu')^2\nu'' - 4\nu^2\nu'\nu''' - 3\nu^2(\nu'')^2 + \nu^3\nu^{(4)} \right]. \quad (198)$$

We now show that  $|r(u)|$  is bounded for  $|\eta| \leq \eta_{\max}$  and  $|u| \leq \delta$ . To this aim, it is useful to introduce the region of the complex plane:

$$\mathcal{A} = \{\operatorname{Re}(z) \in [-\eta_{\max}, \eta_{\max}], \operatorname{Im}(z) \in [-\delta, \delta]\}. \quad (199)$$

First, note that, since all the derivatives of  $\nu(z)$  are analytic (and, hence, continuous) in  $\mathbb{C}$  [46], the term between brackets in (198) is bounded within the bounded and closed set  $\mathcal{A}$ . In addition, for  $z \in \mathcal{A}$ , Eq. (194) implies that  $|\nu(z)| > 0$ . Indeed, if we had  $\nu(z) = 0$  for a certain  $z \in \mathcal{A}$ , then there would exist  $\eta$  and  $u$ , with  $|\eta| \leq \eta_{\max}$  and  $|u| \leq \delta$ , such that  $\nu(\eta + ju) = 0$ , implying:

$$|\hat{\varphi}(u) - 1| = \left| \frac{\nu(\eta + ju)}{\nu(\eta)} e^{-j\psi'(\eta)u} - 1 \right| = 1, \quad (200)$$

and this condition would violate (194). We can accordingly write, for all  $|\eta| \leq \eta_{\max}$  and  $|u| \leq \delta$ :

$$\left| \frac{1}{\nu(\eta + ju)} \right| \leq \max_{z \in \mathcal{A}} \left| \frac{1}{\nu(z)} \right| < \infty, \quad (201)$$

where finiteness follows since  $\nu(z)$  is analytic (and, hence, continuous) in  $\mathbb{C}$ , and it does not vanish within the set  $\mathcal{A}$ . We have in fact shown that  $|r(u)|$  is bounded for  $|\eta| \leq \eta_{\max}$  and  $|u| \leq \delta$ , namely, that a constant  $M = M(\delta, \eta_{\max})$  exists such that  $|r(\bar{u})| \leq M$ . In the light of (196), this completes the proof of (188) and (189).

### B. Technical lemma relevant to Theorem 3

Let us introduce the main quantities involved in the forthcoming derivation.

- The measure transformation:

$$\tilde{m}(dy) = e^{\frac{\theta}{\mu} y - \phi_{k,\mu}(\frac{\theta}{\mu})} m(dy). \quad (202)$$

- The random variable introduced in (148):

$$\mathbf{w}_{k,\mu}^* = \frac{\mathbf{y}_{k,\mu}^* - \phi'_{k,\mu}(\theta/\mu)}{\sqrt{\mu} s_{k,\mu}}, \quad (203)$$

whose cumulative distribution function under the transformed measure has been denoted by  $F_{k,\mu}(w)$ . Using property P3 and applying (186), it is easily seen that the CF of  $\mathbf{w}_{k,\mu}^*$  under the transformed measure is:

$$\tilde{\varphi}_{k,\mu}(u) = \frac{\nu_{k,\mu}\left(\frac{\theta}{\mu} + j\frac{u}{\sqrt{\mu}s_{k,\mu}}\right)}{\nu_{k,\mu}\left(\frac{\theta}{\mu}\right)} e^{-ju\frac{\phi'_{k,\mu}(\theta/\mu)}{\sqrt{\mu}s_{k,\mu}}}, \quad (204)$$

where  $\nu_{k,\mu}(\cdot)$  denotes the MGF (of real or complex argument) of the random variable  $\mathbf{y}_{k,\mu}^*$  under the original measure.

- The normal approximation  $G_{k,\mu}(w)$  defined by (157):

$$G_{k,\mu}(w) = N(w) - \frac{\phi_{k,\mu}'''(\theta/\mu)}{6\mu^{3/2}s_{k,\mu}^3}(w^2 - 1)n(w). \quad (205)$$

The following lemma shows that  $F_{k,\mu}(w)$  converges to  $G_{k,\mu}(w)$  as  $\mu$  goes to zero. From (91) and (147), we conclude that the second term on the RHS of (205) vanishes as  $\mu \rightarrow 0$ . This result will imply that  $\mathbf{w}_{k,\mu}^*$  is asymptotically normal. In addition, regarding  $G_{k,\mu}(w)$  as a higher-order approximation for  $F_{k,\mu}(w)$  (namely, including a correction with respect to the crudest approximation  $N(w)$ ) the lemma provides the useful information that the (worst-case) rate of convergence of the approximation error is in the order of  $\sqrt{\mu}$ . This observation is critical for the proof of Theorem 3.

**LEMMA 3 (Asymptotic normality of  $\mathbf{w}_{k,\mu}^*$  under the transformed measure (202): Estimate of the convergence error).** *Under the same hypotheses of Theorem 3:*

$$\frac{1}{\sqrt{\mu}} \sup_{w \in \mathbb{R}} |F_{k,\mu}(w) - G_{k,\mu}(w)| \xrightarrow{\mu \rightarrow 0} 0. \quad (206)$$

*Proof.* The following classical result [40, page 538] is used in our proof.

*Let  $F$  be a probability distribution with zero mean and characteristic function  $\varphi$ . Suppose that  $F - G$  vanishes at  $\pm\infty$  and that  $G$  has a derivative  $g$  such that  $|g| \leq m$ . Finally, suppose that  $g$  has a continuously differentiable Fourier transform  $\zeta$  such that  $\zeta(0) = 1$  and  $\zeta'(0) = 0$ . Then, for all  $w$  and for all  $T > 0$ :*

$$|F(w) - G(w)| \leq \frac{1}{\pi} \int_{-T}^T \left| \frac{\varphi(u) - \zeta(u)}{u} \right| du + \frac{24m}{\pi T}. \quad (207)$$

It is now immediate to verify that the functions  $F_{k,\mu}(w)$  and  $G_{k,\mu}(w)$  introduced above meet the conditions required for (207) to hold. Indeed,  $G_{k,\mu}(w) \rightarrow 0$  as  $w \rightarrow -\infty$ , and  $G_{k,\mu}(w) \rightarrow 1$  as  $w \rightarrow +\infty$ , implying that the difference  $F_{k,\mu}(w) - G_{k,\mu}(w)$  vanishes at  $\pm\infty$  because  $F_{k,\mu}(w)$  is a cumulative distribution function. Moreover,

$$\begin{aligned} \frac{d}{dw} G_{k,\mu}(w) &= n(w) \left[ 1 - w \frac{\phi_{k,\mu}'''(\theta/\mu)}{3 \mu^{3/2} s_{k,\mu}^3} \right] - \\ &\quad \frac{\phi_{k,\mu}'''(\theta/\mu)}{6 \mu^{3/2} s_{k,\mu}^3} (w^2 - 1) n'(w). \end{aligned} \quad (208)$$

From (91) used with  $r = 3$ , we conclude that the quantity  $\frac{\phi_{k,\mu}'''(\theta/\mu)}{\mu^{3/2} s_{k,\mu}^3}$  vanishes as  $\mu \rightarrow 0$  and, hence, it is bounded for sufficiently small values of  $\mu$ . The required boundedness condition on  $\left| \frac{d}{dw} G_{k,\mu}(w) \right|$  then follows by the boundedness of the functions  $n(w)$ ,  $w n(w)$ , and  $(w^2 - 1)n'(w)$ . Finally, the Fourier transform of (208) is given by [40]:

$$\zeta_{k,\mu}(u) \triangleq e^{-\frac{u^2}{2}} \left[ 1 + \frac{\phi_{k,\mu}'''(\theta/\mu)}{6 \mu^{3/2} s_{k,\mu}^3} (ju)^3 \right]. \quad (209)$$

It is seen that  $\zeta_{k,\mu}(u)$  is continuously differentiable, and that  $\zeta_{k,\mu}(0) = 1$ ,  $\zeta'_{k,\mu}(0) = 0$ .

In order to use the above lemma, let us select, for a given  $\epsilon > 0$ , a constant  $a$  such that:

$$a > \frac{24m}{\epsilon}, \quad (210)$$

and set

$$T = \frac{a}{\sqrt{\mu}} \Rightarrow \frac{24m}{\pi T} \leq \epsilon \sqrt{\mu}. \quad (211)$$

With these particular choices, application of (207) yields:

$$\begin{aligned} &\frac{1}{\sqrt{\mu}} |F_{k,\mu}(w) - G_{k,\mu}(w)| \\ &\leq \frac{1}{\sqrt{\mu}} \int_{-\frac{a}{\sqrt{\mu}}}^{\frac{a}{\sqrt{\mu}}} \left| \frac{\tilde{\varphi}_{k,\mu}(u) - \zeta_{k,\mu}(u)}{u} \right| du + \epsilon. \end{aligned} \quad (212)$$

Due to arbitrariness of  $\epsilon$ , the claim (206) will be proved if we show that the first term on the RHS of (212) vanishes as  $\mu \rightarrow 0$ . To get this result, we split the integral as follows:

$$\frac{1}{\sqrt{\mu}} \int_{-\frac{a}{\sqrt{\mu}}}^{\frac{a}{\sqrt{\mu}}} \left| \frac{\tilde{\varphi}_{k,\mu}(u) - \zeta_{k,\mu}(u)}{u} \right| du = I_1 + I_2, \quad (213)$$

where

$$I_1 = \frac{1}{\sqrt{\mu}} \int_{-\frac{\delta}{\sqrt{\mu}}}^{\frac{\delta}{\sqrt{\mu}}} \left| \frac{\tilde{\varphi}_{k,\mu}(u) - \zeta_{k,\mu}(u)}{u} \right| du, \quad (214)$$

and

$$I_2 = \frac{1}{\sqrt{\mu}} \int_{\frac{\delta}{\sqrt{\mu}} < |u| < \frac{a}{\sqrt{\mu}}} \left| \frac{\tilde{\varphi}_{k,\mu}(u) - \zeta_{k,\mu}(u)}{u} \right| du, \quad (215)$$

for a sufficiently small  $\delta$  whose choice will be carefully addressed in the following. We start by proving that  $I_1$  vanishes as  $\mu \rightarrow 0$ .

*Part i* ( $I_1 \rightarrow 0$  as  $\mu \rightarrow 0$ ). From (88) we know that the MGF (with real argument  $t$ ) of the random variable  $\mathbf{y}_{k,\mu}^*$  under the original measure can be represented as:

$$\nu_{k,\mu}(t) = \prod_{i=1}^{\infty} \prod_{\ell=1}^S \nu(\mu(1-\mu)^{i-1} b_{k,\ell}(i)t), \quad (216)$$

where  $\nu(t) = \mathbb{E}[e^{t\mathbf{x}_k(n)}]$  is the MGF of the local statistic  $\mathbf{x}_k(n)$ . We also know that the MGF  $\nu_{k,\mu}(t)$  can be extended to the complex plane by considering

$$\nu_{k,\mu}(z) = \mathbb{E}[e^{z\mathbf{y}_{k,\mu}^*}] \quad (217)$$

for complex  $z$ . It would be useful to have the same infinite-product representation as in (216) for  $\nu_{k,\mu}(z)$ , namely:

$$\nu_{k,\mu}(z) = \prod_{i=1}^{\infty} \prod_{\ell=1}^S \nu(\mu(1-\mu)^{i-1} b_{k,\ell}(i)z). \quad (218)$$

This turns out to be true, and can be justified by an argument similar to that used in [45], which is now illustrated in detail. Consider the sequence of functions, for  $n = 1, 2, \dots$

$$f_n(z) = \prod_{i=1}^n \prod_{\ell=1}^S \nu(\mu(1-\mu)^{i-1} b_{k,\ell}(i)z). \quad (219)$$

First,  $f_n(z)$  is analytic in  $\mathbb{C}$ , since it consists of the product of functions  $\nu(\cdot)$  that are analytic in  $\mathbb{C}$ . Moreover, by (177), we can write, for  $z = t + ju$ :

$$|f_n(z)| \leq \prod_{i=1}^n \prod_{\ell=1}^S \nu(\mu(1-\mu)^{i-1} b_{k,\ell}(i)t) = f_n(t). \quad (220)$$

For any  $n$ , the RHS is a MGF and, hence, its second derivative over  $t$  is strictly positive, which implies that the maximum of the RHS within an interval  $|t| \leq t_{\max}$  is attained at either one or both endpoints of the interval, namely, that:

$$|f_n(z)| \leq \max\{f_n(-t_{\max}), f_n(t_{\max})\}. \quad (221)$$

But since  $f_n(t)$  converges to  $\nu_{k,\mu}(t) < \infty$  as  $n \rightarrow \infty$ , we have that, for any  $\epsilon > 0$ , a certain  $n_0$  exists such that, for all  $n \geq n_0$ :

$$f_n(-t_{\max}) \leq \nu_{k,\mu}(-t_{\max}) + \epsilon, \quad f_n(t_{\max}) \leq \nu_{k,\mu}(t_{\max}) + \epsilon, \quad (222)$$

which shows that the sequence  $|f_n(z)|$  is uniformly bounded, at least for  $n \geq n_0$ , in the closed strip  $\{z \in \mathbb{C} : |\operatorname{Re}(z)| \leq t_{\max}\}$ . Moreover, since the sequence  $f_n(t)$  ( $t \in \mathbb{R}$ ) admits a limit, we can say that  $f_n(z)$  ( $z \in \mathbb{C}$ ) admits a limit at each point of the real-axis, namely, of the interval  $\{z \in \mathbb{C} : |\operatorname{Re}(z)| \leq t_{\max}, \operatorname{Im}(z) = 0\}$ . This allows invoking Vitali's convergence theorem [49], which asserts that  $f_n(z)$  converges, namely:

$$f_n(z) \xrightarrow{n \rightarrow \infty} f(z) = \prod_{i=1}^{\infty} \prod_{\ell=1}^S \nu(\mu(1-\mu)^{i-1} b_{k,\ell}(i)z), \quad (223)$$

uniformly in each bounded closed subregion of the open strip  $\{z \in \mathbb{C} : |\operatorname{Re}(z)| < t_{\max}\}$ , and the limit function  $f(z)$  is therein analytic. Since we are free to choose  $t_{\max}$ , we conclude that  $f(z)$  is analytic in  $\mathbb{C}$ . Moreover, on the real axis

$$f(t) = \prod_{i=1}^{\infty} \prod_{\ell=1}^S \nu(\mu(1-\mu)^{i-1} b_{k,\ell}(i)t) = \nu_{k,\mu}(t), \quad (224)$$

and since  $\nu_{k,\mu}(z)$  is analytic and equals  $\nu_{k,\mu}(t)$  on the real axis, by [46, Corollary, p. 209] we finally get (218).

In the light of the above finding, we are allowed to use (218) into (204). This yields, after some manipulations:

$$\tilde{\varphi}_{k,\mu}(u) = \prod_{i=1}^{\infty} \prod_{\ell=1}^S \frac{\nu(\eta_{i,\ell} + j u_{i,\ell})}{\nu(\eta_{i,\ell})} e^{-j \psi'(\eta_{i,\ell}) u_{i,\ell}}, \quad (225)$$

where we introduced

$$\eta_{i,\ell} \triangleq (1-\mu)^{i-1} b_{k,\ell}(i) \theta, \quad (226)$$

$$u_{i,\ell} \triangleq (1-\mu)^{i-1} b_{k,\ell}(i) \frac{\sqrt{\mu} u}{s_{k,\mu}}, \quad (227)$$

and used (90) to compute  $\phi'_{k,\mu}(\theta/\mu)$ . Now, from property P4 it is seen that the function

$$\frac{\nu(\eta_{i,\ell} + j u_{i,\ell})}{\nu(\eta_{i,\ell})} e^{-j \psi'(\eta_{i,\ell}) u_{i,\ell}} \quad (228)$$

can be regarded as the CF, evaluated at the point  $u_{i,\ell}$ , of the *shifted* random variable  $\mathbf{x}_\ell(i) - \psi'(\eta_{i,\ell})$  under the measure transformation

$$\tilde{m}(dx) = \frac{e^{\eta_{i,\ell}x}}{\nu(\eta_{i,\ell})} m(dx). \quad (229)$$

We shall denote this CF by  $\hat{\varphi}_{\eta_{i,\ell}}(u_{i,\ell})$ , having made explicit the dependence upon  $\eta_{i,\ell}$ , thus getting:

$$\tilde{\varphi}_{k,\mu}(u) = \prod_{i=1}^{\infty} \prod_{\ell=1}^S \hat{\varphi}_{\eta_{i,\ell}}(u_{i,\ell}). \quad (230)$$

From (227) we get:

$$|u_{i,\ell}| \leq \frac{\sqrt{\mu}|u|}{s_{k,\mu}}. \quad (231)$$

We also know from (147) that  $\lim_{\mu \rightarrow 0} s_{k,\mu} = \sqrt{\phi''(\theta)}$ . By applying the definition of limit, this implies that, for any  $\epsilon > 0$ , and for sufficiently small values of  $\mu$ :

$$|u_{i,\ell}| \leq \sqrt{\mu}|u| \left( \frac{1}{\sqrt{\phi''(\theta)}} + \epsilon \right) \leq \delta \left( \frac{1}{\sqrt{\phi''(\theta)}} + \epsilon \right) \triangleq \bar{\delta}, \quad (232)$$

where the last inequality holds because we are focusing on the integral  $I_1$  in (214) and, hence, we consider the range  $\sqrt{\mu}|u| \leq \delta$ .

Noting now that  $|\eta_{i,\ell}| \leq \theta$  by (226), the choice

$$\bar{\delta}^2 < \frac{2}{\max_{|\eta| \leq \theta} \psi''(\eta)} \quad (233)$$

corresponds to (193), and allows invoking property P4, so that it is legitimate to use the Taylor expansion:

$$\begin{aligned} \log \hat{\varphi}_{\eta_{i,\ell}}(u_{i,\ell}) &= -\psi''(\eta_{i,\ell}) \frac{u_{i,\ell}^2}{2} + \psi'''(\eta_{i,\ell}) \frac{(ju_{i,\ell})^3}{6} + \\ &\quad R(u_{i,\ell}). \end{aligned} \quad (234)$$

This leads to

$$\begin{aligned} \prod_{i=1}^n \prod_{\ell=1}^S \hat{\varphi}_{\eta_{i,\ell}}(u_{i,\ell}) &= \prod_{i=1}^n \prod_{\ell=1}^S e^{\log \hat{\varphi}_{\eta_{i,\ell}}(u_{i,\ell})} \\ &= \prod_{i=1}^n \prod_{\ell=1}^S e^{-\psi''(\eta_{i,\ell}) \frac{u_{i,\ell}^2}{2} + \psi'''(\eta_{i,\ell}) \frac{(ju_{i,\ell})^3}{6} + R(u_{i,\ell})} \\ &= e^{\sum_{i=1}^n \sum_{\ell=1}^S -\psi''(\eta_{i,\ell}) \frac{u_{i,\ell}^2}{2} + \psi'''(\eta_{i,\ell}) \frac{(ju_{i,\ell})^3}{6} + R(u_{i,\ell})}. \end{aligned} \quad (235)$$

We now show that all terms in the exponent are convergent as  $n \rightarrow \infty$ . By using the expressions for  $\eta_{i,\ell}$ ,  $u_{i,\ell}$  in (226) and (227), and recalling the definition of  $\xi_{i,\ell}$  in (87), straightforward manipulations yield:

$$\begin{aligned} \sum_{i=1}^{\infty} \sum_{\ell=1}^S \psi''(\eta_{i,\ell}) \frac{u_{i,\ell}^2}{2} &= \frac{1}{\mu s_{k,\mu}^2} \sum_{i=1}^{\infty} \sum_{\ell=1}^S \xi_{i,\ell}^2 \psi''(\xi_{i,\ell} \theta / \mu) \frac{u^2}{2} \\ &= \frac{\phi''_{k,\mu}(\theta / \mu)}{\mu s_{k,\mu}^2} \frac{u^2}{2} = \frac{u^2}{2}, \end{aligned} \quad (236)$$

having used, in the last two equalities, Eq. (90) and the definition of  $s_{k,\mu}$  in (147). Similarly,

$$\begin{aligned} A_1 &\triangleq \sum_{i=1}^{\infty} \sum_{\ell=1}^S \psi'''(\eta_{i,\ell}) \frac{(ju_{i,\ell})^3}{6} \\ &= \frac{1}{\mu^{3/2} s_{k,\mu}^3} \sum_{i=1}^{\infty} \sum_{\ell=1}^S \xi_{i,\ell}^3 \psi'''(\xi_{i,\ell} \theta / \mu) \frac{(ju)^3}{6} \\ &= \frac{\phi'''_{k,\mu}(\theta / \mu)}{\mu^{3/2} s_{k,\mu}^3} \frac{(ju)^3}{6}. \end{aligned} \quad (237)$$

Finally, by defining

$$A_2 \triangleq \sum_{i=1}^{\infty} \sum_{\ell=1}^S R(u_{i,\ell}), \quad (238)$$

and using (189), we have (recall that  $b_{k,\ell}(i) \leq 1$ ):

$$\begin{aligned} |A_2| &\leq \sum_{i=1}^{\infty} \sum_{\ell=1}^S |R(u_{i,\ell})| \leq \frac{M}{24} \sum_{i=1}^{\infty} \sum_{\ell=1}^S |u_{i,\ell}|^4 \\ &\leq \frac{M}{24 s_{k,\mu}^4} \mu^2 \sum_{i=1}^{\infty} \sum_{\ell=1}^S (1 - \mu)^{4(i-1)} |u|^4 \\ &= \frac{MS}{24 s_{k,\mu}^4} \frac{\mu^2}{1 - (1 - \mu)^4} |u|^4. \end{aligned} \quad (239)$$

We stress that the constant  $M$  appearing in the above formulas bounds uniformly all terms  $|R(u_{i,\ell})|$ . Indeed, as already observed, in the considered range of analysis we have  $|u_{i,\ell}| \leq \bar{\delta}$  and  $|\eta_{i,\ell}| \leq \theta$ . Under this assumption, from property P4 we know that the bounding constant  $M$  depends only on the pair  $(\bar{\delta}, \theta)$ , and, hence, not on the particular pair  $(i, \ell)$ .

Using (230) and (235), it is now legitimate to rewrite

$$\tilde{\varphi}_{k,\mu}(u) = e^{-\frac{u^2}{2} + A_1 + A_2}. \quad (240)$$



Making explicit the definition of  $\zeta_{k,\mu}(u)$  appearing in (214), we have:

$$\begin{aligned}
& |\tilde{\varphi}_{k,\mu}(u) - \zeta_{k,\mu}(u)| \\
&= \left| \tilde{\varphi}_{k,\mu}(u) - e^{-\frac{u^2}{2}} \left( 1 + \frac{\phi_{k,\mu}'''(\theta/\mu)}{6 \mu^{3/2} s_{k,\mu}^3} (ju)^3 \right) \right| \\
&= e^{-\frac{u^2}{2}} \left| \tilde{\varphi}_{k,\mu}(u) e^{\frac{u^2}{2}} - 1 - \frac{\phi_{k,\mu}'''(\theta/\mu)}{\mu^{3/2} s_{k,\mu}^3} \frac{(ju)^3}{6} \right| \\
&= e^{-\frac{u^2}{2}} |e^{A_1+A_2} - 1 - A_1| \\
&\leq e^{-\frac{u^2}{2}} \left( |A_2| + \frac{|A_2|^2}{2} \right) e^{|A_1|+|A_2|}, \tag{241}
\end{aligned}$$

where, in the last inequality, we resorted to the following result, which is known to hold for any pair of complex numbers  $z$  and  $w$  — see, e.g., [40, Eq. (2.8), p. 534]:

$$|e^z - 1 - w| \leq \left( |z - w| + \frac{1}{2}|w|^2 \right) e^{\max(|z|, |w|)}. \tag{242}$$

Consider first the term  $A_1$ . Using (147) and (91) with  $r = 3$ , we can write:

$$\lim_{\mu \rightarrow 0} \left| \frac{\phi_{k,\mu}'''(\theta/\mu)}{\mu^2 s_{k,\mu}^3} \right| = \left| \frac{\phi'''(\theta)}{[\phi''(\theta)]^{3/2}} \right|. \tag{243}$$

Making explicit the definition of limit, for any  $\epsilon > 0$ , and for  $\mu$  small enough, it holds that:

$$\begin{aligned}
|A_1| &= \frac{1}{6\mu} \left| \frac{\phi_{k,\mu}'''(\theta/\mu)}{\mu^2 s_{k,\mu}^3} \right| |\sqrt{\mu}u|^3 \\
&\leq \frac{1}{6\mu} \left( \left| \frac{\phi'''(\theta)}{[\phi''(\theta)]^{3/2}} \right| + \epsilon \right) |\sqrt{\mu}u|^3 \\
&= \frac{1}{\mu} a_1 |\sqrt{\mu}u|^3, \tag{244}
\end{aligned}$$

where the constant  $a_1$  has been implicitly defined. We switch to the term  $A_2$ . Applying L'Hospital's rule [43], and using again (147), we see that:

$$\lim_{\mu \rightarrow 0} \left( \frac{\mu}{1 - (1 - \mu)^4} \frac{1}{s_{k,\mu}^4} \right) = \frac{1}{4[\phi''(\theta)]^2}, \tag{245}$$

implying that, for any  $\epsilon > 0$ , and for sufficiently small values of  $\mu$ :

$$\begin{aligned}
|A_2| &\leq \frac{1}{\mu} \frac{MS}{24 s_{k,\mu}^4} \frac{\mu}{1 - (1 - \mu)^4} (\sqrt{\mu}u)^4 \\
&\leq \frac{1}{\mu} \frac{MS}{24} \left( \frac{1}{4[\phi''(\theta)]^2} + \epsilon \right) (\sqrt{\mu}u)^4 \\
&= \frac{1}{\mu} a_2 (\sqrt{\mu}u)^4, \tag{246}
\end{aligned}$$

where the constant  $a_2$  has been implicitly defined. Observe now that, for any  $a > 0$ , any  $0 \leq x \leq \delta$ , and for any integers  $m$  and  $n$  such that  $m < n$ :

$$\delta^{n-m} \leq 1/a \Rightarrow ax^n \leq x^m. \quad (247)$$

Setting  $x = |\sqrt{\mu}u|$ , and recalling that in the considered range  $|\sqrt{\mu}u| \leq \delta$ , the above relationship implies that the parameter  $\delta$  can be certainly chosen to satisfy:

$$a_1 |\sqrt{\mu}u|^3 \leq \frac{|\sqrt{\mu}u|^2}{8}, \quad a_2 (\sqrt{\mu}u)^4 \leq \frac{|\sqrt{\mu}u|^2}{8}, \quad (248)$$

yielding (in the considered range):

$$|A_1| \leq \frac{u^2}{8}, \quad |A_2| \leq \frac{u^2}{8}. \quad (249)$$

Finally, using these inequalities in (241), the integrand in (214) can be upper bounded as:

$$\frac{1}{\sqrt{\mu}} \left| \frac{\tilde{\varphi}_{k,\mu}(u) - \zeta_{k,\mu}(u)}{u} \right| \leq e^{-\frac{u^2}{4}} \left( \sqrt{\mu} a_2 |u|^3 + \mu^{3/2} \frac{a_2^2 |u|^7}{2} \right), \quad (250)$$

implying that  $I_1 \rightarrow 0$  as  $\mu \rightarrow 0$ .

*Part ii* ( $I_2 \rightarrow 0$  as  $\mu \rightarrow 0$ ). The following chain of inequalities holds true:

$$\begin{aligned} I_2 &\leq \frac{1}{\sqrt{\mu}} \int_{\frac{\delta}{\sqrt{\mu}} < |u| < \frac{a}{\sqrt{\mu}}} \left| \frac{\tilde{\varphi}_{k,\mu}(u)}{u} \right| du + \\ &\quad \frac{1}{\sqrt{\mu}} \int_{|u| > \frac{\delta}{\sqrt{\mu}}} e^{-\frac{u^2}{2}} \left| \frac{1 + \frac{\phi_{k,\mu}'''(\theta/\mu)}{6\mu^{3/2}s_{k,\mu}^3} (ju)^3}{u} \right| du \\ &\leq \frac{1}{\delta} \int_{\frac{\delta}{\sqrt{\mu}} < |u| < \frac{a}{\sqrt{\mu}}} |\tilde{\varphi}_{k,\mu}(u)| du + \\ &\quad \frac{1}{\delta} \int_{|u| > \frac{\delta}{\sqrt{\mu}}} e^{-\frac{u^2}{2}} du + \\ &\quad \frac{|\phi_{k,\mu}'''(\theta/\mu)|}{6\mu^2 s_{k,\mu}^3} \int_{|u| > \frac{\delta}{\sqrt{\mu}}} e^{-\frac{u^2}{2}} u^2 du, \end{aligned} \quad (251)$$

which follows by repeated application of the triangle inequality for complex numbers  $|z + w| \leq |z| + |w|$ . Since the integrals

$$\int_{-\infty}^{\infty} e^{-\frac{u^2}{2}} du, \quad \int_{-\infty}^{\infty} e^{-\frac{u^2}{2}} u^2 du \quad (252)$$

are convergent, and since the quantities  $\phi_{k,\mu}'''(\theta/\mu)/\mu^2$  and  $s_{k,\mu}$  converge in view of (91) and (147), the last two integrals on the RHS of (251) vanish as  $\mu \rightarrow 0$ .

It remains to prove that the first integral on the RHS of (251) vanishes as  $\mu \rightarrow 0$ . To this aim, observe that, by (225):

$$|\tilde{\varphi}_{k,\mu}(u)| = \prod_{i=1}^{\infty} \prod_{\ell=1}^S \frac{|\nu(\eta_{i,\ell} + ju_{i,\ell})|}{\nu(\eta_{i,\ell})}. \quad (253)$$

Since, from (177), we know that

$$\frac{|\nu(\eta_{i,\ell} + ju_{i,\ell})|}{\nu(\eta_{i,\ell})} \leq 1, \quad (254)$$

all the terms of the product in (253) are not greater than one, implying that, in particular:

$$|\tilde{\varphi}_{k,\mu}(u)| \leq \prod_{i=\lceil 1/\mu \rceil}^{2\lfloor 1/\mu \rfloor} \prod_{\ell=1}^S \frac{|\nu(\eta_{i,\ell} + ju_{i,\ell})|}{\nu(\eta_{i,\ell})}, \quad (255)$$

where  $\lceil x \rceil$  (resp.,  $\lfloor x \rfloor$ ) denotes the smallest (resp., the largest) integer not smaller (resp., not greater) than  $x$ . Since in the above product we have  $1/\mu \leq \lceil 1/\mu \rceil \leq i \leq 2\lfloor 1/\mu \rfloor \leq 2/\mu$ , and since, in the range of interest  $\delta < |\sqrt{\mu}u| < a$ , for  $u_{i,\ell}$  defined in (227) we can write:

$$\frac{(1-\mu)^{2/\mu-1} b_{k,\ell}(i)}{s_{k,\mu}} \delta < |u_{i,\ell}| < \frac{(1-\mu)^{1/\mu-1} b_{k,\ell}(i)}{s_{k,\mu}} a. \quad (256)$$

The following convergences hold:

$$\lim_{\mu \rightarrow 0} (1-\mu)^{1/\mu} = 1/e, \quad [\text{known limit}] \quad (257)$$

$$\lim_{i \rightarrow \infty} b_{k,\ell}(i) = p_\ell, \quad [\text{Eq. (18)}] \quad (258)$$

$$\lim_{\mu \rightarrow 0} s_{k,\mu}(i) = \sqrt{\phi''(\theta)}. \quad [\text{Eq. (147)}] \quad (259)$$

Using the explicit definition of limit, the above three relationships imply that, for any  $\epsilon > 0$ , it is possible to choose  $\mu$  sufficiently small so as to ensure that, in the considered range  $i \geq \lceil 1/\mu \rceil$ :

$$v_1 \triangleq \left( \frac{p_\ell/e^2}{\sqrt{\phi''(\theta)}} - \epsilon \right) \delta < |u_{i,\ell}| < \left( \frac{p_\ell/e}{\sqrt{\phi''(\theta)}} + \epsilon \right) a \triangleq v_2, \quad (260)$$

where, due to arbitrariness of  $\epsilon$ , we can safely assume that  $v_1 > 0$ .

Moreover, for  $\eta_{i,\ell}$  defined in (226), we have  $0 \leq \eta_{i,\ell} \leq \theta$ . Accordingly, by defining the set:

$$\mathcal{S}_\ell = \{(\eta, u) \in \mathbb{R}^2 : 0 \leq \eta \leq \theta, v_1 \leq |u| \leq v_2\}, \quad (261)$$

we have:

$$\int_{\frac{\delta}{\sqrt{\mu}} < |u| < \frac{a}{\sqrt{\mu}}} |\tilde{\varphi}_{k,\mu}(u)| du \leq 2 \frac{a-\delta}{\sqrt{\mu}} c^{1/\mu+1}, \quad (262)$$

where

$$\begin{aligned} c &= \prod_{\ell=1}^S \left( \max_{(\eta, u) \in \mathcal{S}_\ell} \frac{|\nu(\eta + ju)|}{\nu(\eta)} \right) \\ &= \prod_{\ell=1}^S \frac{|\nu(\bar{\eta}_\ell + j\bar{u}_\ell)|}{\nu(\bar{\eta}_\ell)}. \end{aligned} \quad (263)$$

In the above equation,  $(\bar{\eta}_\ell, \bar{u}_\ell)$  are the pairs where the maxima are attained, which belong to  $\mathcal{S}_\ell$  since  $\frac{|\nu(\eta + ju)|}{\nu(\eta)}$ , regarded as a function of the pair  $(\eta, u)$ , is continuous.

In view of property P3,  $\frac{\nu(\eta + ju)}{\nu(\eta)}$  is (as a function of  $u$ ) nothing but the CF of  $\mathbf{x}_k(n)$  under an exponential measure translation. If  $\mathbf{x}_k(n)$  is not lattice [40], the magnitude of the CF attains the value one only at the origin. This implies that all the maxima appearing in (263) are strictly less than 1, because  $u$  is bounded away from zero in  $\mathcal{S}_\ell$ . Since, for any  $c < 1$ , the function  $c^{1/\mu}$  vanishes faster than any power of  $\mu$  as  $\mu \rightarrow 0$ , we conclude that the integral in (262) vanishes as  $\mu$  goes to zero, and the proof of the lemma is complete.  $\square$

REMARK VI. The proof of Lemma 3 assumes that the local statistic  $\mathbf{x}_k(n)$  is not lattice. The distribution of a lattice random variable with span  $\Delta$  is concentrated at the points  $d, d \pm \Delta, d \pm 2\Delta, \dots$  for a certain real number  $d$ . The pertinent CF becomes:

$$\frac{\nu(\eta + ju)}{\nu(\eta)} = \frac{\sum_{m=-\infty}^{\infty} p_m e^{\eta(d+m\Delta)} e^{jtd(m\Delta)}}{\nu(\eta)}, \quad (264)$$

where  $p_m = \mathbb{P}[\mathbf{x}_k(n) = d + m\Delta]$ . Accordingly, we have

$$\frac{|\nu(\eta + jt)|}{\nu(\eta)} = \frac{\left| \sum_{m=-\infty}^{\infty} p_m e^{\eta(d+m\Delta)} e^{jtm\Delta} \right|}{\nu(\eta)}, \quad (265)$$

which shows that the magnitude of the CF has period  $2\pi/\Delta$  and, in particular, it assumes the value one at all points  $t = 2\pi h/\Delta$ , for  $h = 0, \pm 1, \pm 2, \dots$ , which would violate the last step in our proof above.  $\square$

## REFERENCES

- [1] P. Braca, S. Marano, V. Matta, and A. H. Sayed, "Asymptotic performance of adaptive distributed detection over networks," *submitted for publication*. Available as arXiv:1401.5742v2 [cs.IT], Jan. 2014.
- [2] P. Braca, S. Marano, V. Matta, and A. H. Sayed, "Large deviations analysis of adaptive distributed detection," in *Proc. IEEE ICASSP*, Florence, Italy, May 2014, pp. 6112–6116.
- [3] V. Matta, P. Braca, S. Marano, and A. H. Sayed, "Exact asymptotics of distributed detection over adaptive networks," in *Proc. IEEE ICASSP*, Brisbane, Australia, Apr. 2015, pp. 3377–3381.
- [4] C. G. Lopes and A. H. Sayed, "Diffusion least-mean squares over adaptive networks: Formulation and performance analysis," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3122–3136, Jul. 2008.
- [5] F. S. Cattivelli and A. H. Sayed, "Diffusion LMS strategies for distributed estimation," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1035–1048, Mar. 2010.
- [6] A. H. Sayed, S.-Y. Tu, J. Chen, X. Zhao, and Z. J. Towfic, "Diffusion strategies for adaptation and learning over networks," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 155–171, May 2013.
- [7] A. H. Sayed, "Adaptive networks," *Proceedings of the IEEE*, vol. 102, no. 4, pp. 460–497, Apr. 2014.
- [8] P. Braca, S. Marano, and V. Matta, "Enforcing consensus while monitoring the environment in wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3375–3380, Jul. 2008.
- [9] P. Braca, S. Marano, V. Matta, and P. Willett, "Asymptotic optimality of running consensus in testing statistical hypotheses," *IEEE Trans. Signal Process.*, vol. 58, no. 2, pp. 814–825, Feb. 2010.
- [10] —, "Consensus-based Page's test in sensor networks," *Signal Processing*, vol. 91, no. 4, pp. 919–930, Apr. 2011.
- [11] D. Bajovic, D. Jakovetic, J. Xavier, B. Sinopoli, and J. M. F. Moura, "Distributed detection via Gaussian running consensus: Large deviations asymptotic analysis," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4381–4396, Sep. 2011.
- [12] D. Bajovic, D. Jakovetic, J. M. F. Moura, J. Xavier, and B. Sinopoli, "Large deviations performance of consensus+innovations distributed detection with non-Gaussian observations," *IEEE Trans. Signal Process.*, vol. 60, no. 11, pp. 5987–6002, Nov. 2012.
- [13] D. Jakovetic, J. M. F. Moura, and J. Xavier, "Distributed detection over noisy networks: Large deviations analysis," *IEEE Trans. Signal Process.*, vol. 60, no. 8, pp. 4306–4320, Aug. 2012.
- [14] J. Tsitsiklis, D. Bertsekas, and M. Athans, "Distributed asynchronous deterministic and stochastic gradient optimization algorithms," *IEEE Trans. Autom. Control*, vol. 31, no. 9, pp. 803–812, Sep. 1986.
- [15] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Systems and Control Letters*, vol. 53, no. 1, pp. 65–78, Sep. 2004.
- [16] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2508–2530, Jun. 2006.
- [17] A. Nedic and A. Ozdaglar, "Distributed subgradient methods for multiagent optimization," *IEEE Trans. Autom. Control*, vol. 54, no. 1, pp. 48–61, Jan. 2009.
- [18] A. G. Dimakis, S. Kar, J. M. F. Moura, M. G. Rabbat, and A. Scaglione, "Gossip algorithms for distributed signal processing," in *Proc. IEEE*, vol. 98, no. 11, pp. 1847–1864, Nov. 2010.
- [19] S. Kar and J. M. F. Moura, "Convergence rate analysis of distributed gossip (linear parameter) estimation: Fundamental limits and tradeoffs," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 4, pp. 674–690, Aug. 2011.
- [20] M. Longo, T. D. Lookabaugh, and R. M. Gray, "Quantization for decentralized hypothesis testing under communication constraints," *IEEE Trans. Inf. Theory*, vol. 36, no. 2, pp. 241–255, Mar. 1990.

- [21] P. K. Varshney, *Distributed Detection and Data Fusion*. Springer-Verlag, New York, 1997.
- [22] R. Viswanathan and P. K. Varshney, "Distributed detection with multiple sensors: Part I—Fundamentals," in *Proc. IEEE*, vol. 85, no. 1, pp. 54–63, Jan. 1997.
- [23] R. S. Blum, S. A. Kassam, and H. V. Poor, "Distributed detection with multiple sensors: Part II—Advanced topics," in *Proc. IEEE*, vol. 85, no. 1, pp. 64–79, Jan. 1997.
- [24] J. F. Chamberland and V. V. Veeravalli, "Decentralized detection in sensor networks," *IEEE Trans. Signal Process.*, vol. 51, no. 2, pp. 407–416, Feb. 2003.
- [25] J. F. Chamberland and V. V. Veeravalli, "Wireless sensors in distributed detection applications," *IEEE Signal Process. Mag.*, vol. 24, no. 3, pp. 16–25, May 2007.
- [26] B. Chen, L. Tong, and P. K. Varshney, "Channel-aware distributed detection in wireless sensor networks," *IEEE Signal Process. Mag.*, vol. 23, no. 4, pp. 16–26, Jul. 2006.
- [27] V. Saligrama, M. Alanyali, and O. Savas, "Distributed detection in sensor networks with packet losses and finite capacity links," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4118–4132, Nov. 2006.
- [28] J. N. Tsitsiklis, "Decentralized detection," *Adv. Statistical Signal Processing*, vol. 2, pp. 297–344, 1993.
- [29] P. Willett, P. F. Swaszek, and R. S. Blum, "The good, bad and ugly: Distributed detection of a known signal in dependent Gaussian noise," *IEEE Trans. Signal Process.*, vol. 48, no. 12, pp. 3266–3279, Dec. 2000.
- [30] K. Liu and A. M. Sayeed, "Asymptotically optimal decentralized type-based detection in wireless sensor networks," in *Proc. IEEE ICASSP*, Montreal, Canada, May 2004, vol. 3, pp. 873–876.
- [31] S. Barbarossa, S. Sardellitti, and P. Di Lorenzo, "Distributed detection and estimation in wireless sensor networks," in *Academic Press Library in Signal Processing*, vol. 2, R. Chellapa and S. Theodoridis, Eds., pp. 329–408, Academic Press, Elsevier, 2014.
- [32] F. S. Cattivelli and A. H. Sayed, "Distributed detection over adaptive networks using diffusion adaptation," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 1917–1932, May 2011.
- [33] R. R. Bahadur, and R. R. Rao, "On deviations of the sample mean," *The Annals of Mathematical Statistics*, vol. 31, no. 4, pp. 1015–1027, Dec. 1960.
- [34] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*. Springer, 1998.
- [35] F. den Hollander, *Large Deviations*. American Mathematical Society, 2008.
- [36] R. Horn and C. Johnson, *Matrix Analysis*. Cambridge University Press, Cambridge, UK, 1985.
- [37] A. H. Sayed, *Adaptive Filters*. Wiley, NJ, 2008.
- [38] H. V. Poor, *An Introduction to Signal Detection and Estimation*. Springer-Verlag, 1988.
- [39] S. A. Kassam, *Signal Detection in Non-Gaussian Noise*. Springer-Verlag, 1987.
- [40] W. Feller, *An Introduction to Probability and Its Applications*, vol. 2, Wiley, NY, 1971.
- [41] H. Cramér, "Sur un nouveau théorème-limite de la théorie des probabilités," in: *Actualités Scientifiques et Industrielles*, Hermann, Paris, no. 736, pp. 5–23, 1938.
- [42] J. A. Bucklew, *Large Deviations Techniques in Decision, Simulation and Estimation*. Wiley, 1990.
- [43] W. Rudin, *Principles of Mathematical Analysis*. McGraw-Hill, 1976.
- [44] H. L. Van Trees, *Detection, Estimation, and Modulation Theory. Part I*. Wiley, 1968, (reprinted, 2001).
- [45] J. H. Curtiss, "A note on the theory of moment generating functions," *The Annals of Statistics*, vol. 13, no. 4, pp. 430–433, 1942.
- [46] W. Rudin, *Real and Complex Analysis*. McGraw-Hill, 1987.

- [47] H. Shao, *Mathematical Statistics*, 2nd ed., Springer, 2003.
- [48] R. N. Bhattacharya and R. R. Rao, *Normal Approximation and Asymptotic Expansions*. Robert E. Krieger Publishing Company, 1986.
- [49] E. C. Titchmarsh, *The Theory of Functions*. Oxford University Press, 1939.